DOCUMENT RESUME

ED 026 966 24 HE 000 655

By-Dubin, Robert; Taveggia, Thomas C.
The Teaching-Learning Paradox: A Comparative Analysis of College Teaching Methods.
Oregon Univ., Eugene. Center for Advanced Study of Educational Administration.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No-BR-5-0217
Pub Date Dec 68
Contract-OEC-4-10-163
Note-85p.
EDRS Price MF-$0.50 HC-$4.35
Descriptors-*Comparative Analysis, *Higher Education, *Learning, *Research, *Teaching Methods, Teaching Models, Tests

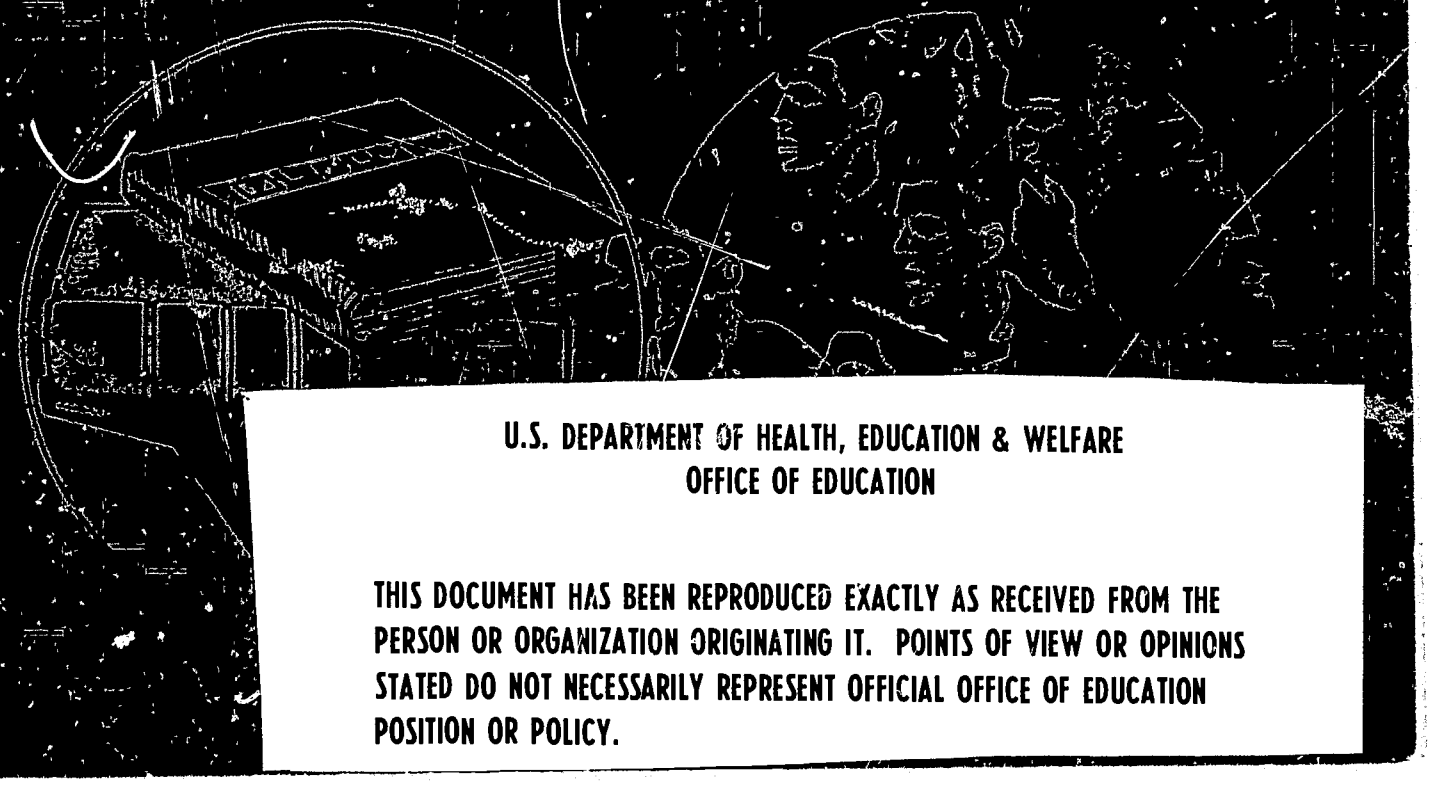
The results of a careful and systematic reanalysis of the data of almost 100 comparative studies of different college teaching methods are reported. No shred of evidence was found to indicate any basis for preferring one teaching method over another as measured by the performance of students on course examinations. Underlying all theories concerning the efficacy of one teaching method over another is an implicit model of how teaching and learning are linked. However, we really do not know what the linkage is. The need for establishing clear and unequivocal links between a theory of learning and a theory of teaching is a vital one. To answer this need, research on comparative college teaching methods has to move in new directions: (1) to find the commonalities among all distinctive college teaching methods; and (2) to develop models of the teaching-learning situation. Nothing new will be discovered about college teaching methods until new questions are asked and their answers sought in research that is significantly different from that pursued in the past. The Appendices explain how conclusions were reached. A Bibliography is included. (Author/JS)

ROBERT DUBIN

THOMAS C. TAVEGGIA

TEACHING-LEARNING Paradox

A Comparative Analysis of College Teaching Methods



Bur. No. 5-0217
Proj. & Project No. 5-0217-4-5
Funding Authority:
Cooperative Research,
PL 83-531
as amended
Contract No. 4-10-163

The Teaching-Learning Paradox

The Teaching-Learning Paradox

A Comparative Analysis of College Teaching Methods

Robert Dubin
Thomas C. Taveggia

Center for the Advanced Study of Educational Administration
University of Oregon Eugene, Oregon
1968

The Center for the Advanced Study of Educational Administration at the University of Oregon is a national research and development center which was established under the provisions of the Cooperative Research Program of the United States Office of Education. The research and development reported herein was performed pursuant to a contract with the Division of Educational Laboratories of the Office of Education. No federal funds were used in the publication of this report.

Printed in the United States of America at the
University of Oregon Press, Eugene, Oregon
Designed and edited by Joanne M. Kitchel
The text of this book is set in Bodoni type
on Simpson's eggshell text
LIBRARY OF CONGRESS CATALOG NUMBER 68-66809

The activities of teaching, learning and examining in higher education have received remarkably little attention from experimentalists. The suggestion that belief, however hallowed, stands in need of empirical support is often met with incredulous opposition, and men whose academic work is based upon a rigorous testing of accepted ideas, tend to think it somehow indecent to apply the same standard of inquiry to their own teaching and examing practices. Few university teachers are even aware that many of their instructional problems have already been investigated experimentally, and only a tiny minority take the trouble to acquaint themselves with the results.

J. P. POWELL

"Experimentation and Teaching in Higher Education," *Educational Research*, 6, (1964), p. 179.

Foreword

This is a polemical tract in the best sense of the term. We feel that the proper use of scientific knowledge is (1) to be the foundation for future additions to knowledge, and (2) to be the knowledge base for policy in the practical affairs of men.

We do not hesitate to demand that future research on college teaching methods be guided by new theory. The old ideas about pedagogy at the college level are simply wrong. We have assembled in this monograph the data base to make this conclusion incontrovertible. We, therefore, adopt the stance that it is our responsibility to challenge researchers on comparative college teaching methods to start anew from better theory. Any further replication of the four decades of research here reanalyzed has an almost zero probability of changing the conclusions that speak so loudly from these data.

We are equally unhesitant in suggesting that policy-makers who decide about college teaching methods either use their prejudices as a basis for decision (which will produce policy no better than that grounded on other people's prejudices), or feel free to determine policy decisions on grounds other than allegations about pedagogy and learning (e.g. cost, space, time, convenience criteria).

So, we are polemical in charging our research colleagues to make a marked shift in the strategy of doing their research on pedagogy; and in urging academic policy-makers to broaden the grounds for their decisions on college teaching methods.

We think the weight of our conclusions have the special merit of being data-based. We are not polemical by arguing a preferred position against all alternatives. We turned to a voluminous research literature on com-

parative college teaching methods and asked: "How do the facts add up?" The answers are astonishingly clear, (indeed almost unique among behavioral science data in this respect): We cannot claim superiority for any among different teaching methods used to convey subject content to the student.

ROBERT DUBIN
THOMAS C. TAVEGGIA

September, 1968
UNIVERSITY OF OREGON, Eugene, Oregon

Contents

Illustrations

CHAPTER
1

The Issues Posed

It is an important part of the folklore of college teaching that certain methods of instruction are preferred over others. The most preferred methods probably vary by discipline as well as temperament of the professor. Just as important as having preferences is the supporting belief or belief-system that alleges superiority for the preferred teaching method.

We, therefore, enter into an area of strong affect and little concern for facts in trying to assess the relative utility of various college teaching methods. Any conclusion we may draw regarding the facts comparing consequences of various college teaching methods are likely to be challenged to the extent that they violate the belief-system of a given group of professors or college administrators. The result may very well be that we will be forced to argue about the viability and desirability of given belief-systems and not data, or else we will have to acknowledge defeat in the face of the reader's conclusion: "Don't bother me with the facts, my mind is already made up."

These are times of vast change in educational performance, philosophy, and goals. It is a desirable assumption, however shaky its accuracy, to expect that the choice of changed directions will be guided by a significant knowledge base. Insofar as college teaching methods may be subject to change, we would hold that the direction the change takes should be guided by knowledge about the comparative advantage of a given method

of instruction. To this end we have marshaled all the available *data* from comparative studies of college teaching methods.

We have adopted a fundamental procedure, practiced by the senior author, of adding up the *data* of a field, rather than the *conclusions* of researchers who have marshaled the data. In this respect we have made a unique contribution, for all the previous summaries of research in the area of comparative college teaching methods have summarized authors' conclusions, and have not gone back to the basic data from which these conclusions derive.

In cumulating the data from studies of comparative college teaching methods we are following the time-honored procedures of the natural sciences where it has long been the practice to add up the data on a subject as a basis for testing models of the phenomenon, or to build new models on the basis of the accumulated data.

Teaching versus Learning

Analytically we want to measure the utility of one college teaching method over another. We believe that teaching is a technology, the content of which can be rationally ordered by some distinctive model of the teaching process. Thus, the lecture method assumes the superior knowledge of the lecturer and therefore places in his hands the selection of subject matter to be covered, the depth of coverage to be employed, the balance between content and illustration, the length of the lecture period, and the frequency of the lectures in a given period of time. The authoritarian image of the lecturer which derives from this model of the lecture teaching technology does not influence in any way the consistency of the conclusions which derive from the assumption of superior knowledge and information on the part of the lecturer. By way of contrast, self-study (this usually means a reading list and a limited period of time to complete the reading) limits the superiority of the instructor to a knowledge of the relevant bibliography of his field. Beyond that it is assumed that the student learns through interaction with printed materials. The book rather than the instructor becomes the teacher.

It seems reasonable to assume that given such distinctive teaching technologies as lecture, on the one hand, and self-study, on the other hand, there should be measurable differences in outcomes of these two methods. It is the very reasonableness of such an expectation that leads to the conclusion that there will, indeed, be measurable differences between any two contrasting teaching methods.

Underlying all belief-systems in the efficacy of one teaching method

over another is an implicit model of how teaching and learning are linked. It is invariably assumed that when a teacher is teaching a student, the student is learning because of the teacher's intervention. This assumed linkage between teaching and learning is the morass in which much of the controversy about various teaching methods bogs down. For while the apparent focus of discussion is upon the alleged advantage of one teaching method over another, the outcome measured is usually declared to be some aspect of student learning. If it can be shown that learning of content is not differentiated between two or more teaching methods, then it may be alleged that: "The student learns a point of view;" "The student learns to think;" "The student learns sensitivity to others;" "The student achieves inspiration;" "The student develops a sense of relevance;" or countless other variations on these themes, all of which center on allegations regarding what the student learns most effectively from a preferred teaching method.

It has only been in the current decade that recognition has grown apace that we really do not know what the linkage is between teaching and learning. Gage and Hilgard have been leaders in alerting the schoolmen to the need for establishing the clear and unequivocal links between a theory of learning and a theory of teaching. We subscribe wholeheartedly to this position. This monograph will have made a significant contribution if it does nothing more than fortify the conclusion that we have not yet established adequate theories of the linkages between teaching and learning.

Measuring the Results of Teaching

The comparative study of college teaching methods has largely focused attention on measuring the results of college teaching through a final examination given to the students. Examinations of this sort are typically content-oriented, and are designed to determine how much of the content presented by the teacher can be recalled, after some delay, during the final examination of the student.

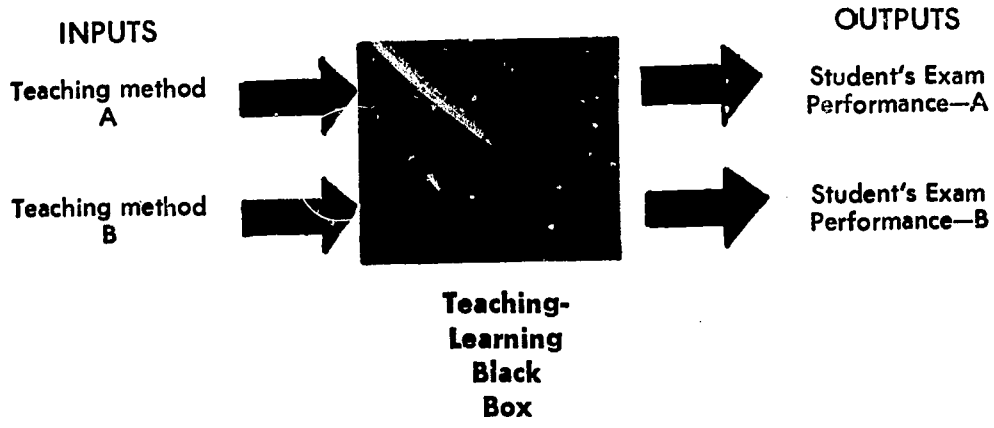
The problem with which we are confronted is simply stated. Does the final examination in a course measure the learning of the student? Does the final examination in a course measure the teaching methods of the teacher? Or does the final examination in a course measure some combination of learning and teaching? We are in no position to give a positive answer to any of these questions.

Most of the time we will be measuring whatever happened by the end of a course, utilizing a final examination and the grade assigned to it.

In rare cases we will have an opportunity to measure gains made by a student, from some measured level of knowledge before taking a course, to the measured level of knowledge determined by examination at the end of a course of instruction. In a few instances we have studies that evaluate other outcomes measured on the student body.

We must, therefore, start with the difficult issue of deciding what, if anything, is being measured by the outcomes employed in the numerous studies of comparative college teaching methods. The easiest way to conceptualize this is to employ the familiar black box analogy. The inputs into a black box labeled "teaching-learning" situation are two or more distinctive teaching technologies covering the same subject matter and

THE "TEACHING-LEARNING" BLACK BOX



subject matter area. Whatever happened inside the black box, at the end of a prescribed period of time known as the length of the course, an output is measured on the student body by a final examination. Now, without making any assumptions about what goes on inside the teaching-learning black box, we may simply conclude that one group of students subject to teaching method A, when compared with an equivalent group of students subject to teaching method B, will produce either the same output or a distinctly different output in the form of group mean examination scores.

If the measured outputs differ by a statistically significant amount we may then make one of two assumptions: (a) that the differential inputs represented by teaching methods are related to the outputs from the teaching-learning situation; or (b) that the measuring instrument known as a final examination is so unstable as to produce random differences of a magnitude greater than zero.

If we make the second assumption we then focus on the measurement

of outcome and contend that examinations are so inadequate that their vagaries are sufficient to account for major differences in measured group outputs. This assumption about the imperfections of examinations has significant implications for the administration of students in colleges and universities. It calls into question the entire student grading system, the system for advancement to a higher level, the basis for making scholarship and fellowship awards, and the retention of students for a normal academic cycle. Thus, to accept the second assumption is to call into question far more than merely the findings on comparative college teaching methods. We are not, however, asserting that because final examinations have become so highly institutionalized that this is sufficient to justify their existence, or their utilization to measure anything. We merely call your attention to some of the consequences of discrediting final examinations as measuring anything accurately.

We start with the first assumption: that the differential inputs represented by two or more teaching methods are related to the outputs from the teaching-learning situation. We do believe that the examinations used to measure the output of the teaching-learning black box are sufficiently standardized in the given comparative situation that any differences measured by this method, if statistically significant, may not be ascribed to the variability in the measuring instrument itself. Indeed, in most of the comparative studies, the same final examination is used for the groups compared, so that the results are influenced equally by the character and quality of the examinations used to measure student performance.

If we look at the bulk of our findings and discover, as we will, that there are relatively few significant differences among the various teaching methods as measured by examinations, then one might argue, ingeniously, that whatever the examinations measure they do not measure the distinctive and unique inputs of compared teaching methods. This, of course, has been the principal argument of those who profess surprise whenever the teaching method they prefer does not turn out to be superior to one with which it is compared in a given study. This line of argument may be very significant and it is indeed one of the principal conclusions to which we come in CHAPTER 4. There we will point out as one of several recommendations that any strategy of continued studies of comparative college teaching methods designed to produce a significant payoff must proceed to examine outcomes of the teaching-learning black box other than student performance on final examinations. We will not argue, however, that because we find no differences between two teaching methods that there must be such differences if we will only search hard enough for

them. We prefer to leave the issue of whether differences really exist to the future research for resolution.

The Teaching-Learning Black Box

The fundamental conclusion of this study is that we can no longer afford to consider the teaching-learning black box as an unexplored realm. As already pointed out, Gage and Hilgard along with others have, during the current decade, raised questions about the appropriateness of ignoring the linkage between teaching and learning. We would now add to this that the state-of-the-art analysis represented by this monograph leads to an unequivocal conclusion that the most important analytical problem faced is to explore the content of the teaching-learning black box.

We can no longer assume either a traditional or a comfortable belief about the linkages between teaching and learning. We simply do not know what these linkages are and it is high time that we explore them.

The urgency of the need to explore the teaching-learning connection is emphasized by recalling to you that all the studies here analyzed deal with adult learning. We have limited our attention to a population of college-level students. These are typically 18 years of age and older and the average age is well above 20. There is a high probability that theories or models of teaching-learning appropriate to the young may not be appropriate to near-adults or adults.

For one thing it is very clear that college students have an experience base as well as a volume of learned knowledge that far exceeds that of youngsters. This must surely have an impact upon the kind of teaching that would be particularly appropriate to learning at that age and experience level.

The college student is typically an active participant in the selection of subject areas of his own learning. This selection process is enforced by the institutional requirements of declaring a major and pursuing prescribed courses of instruction to fulfill the major and/or minor requirements for a degree. However minimal the amount of voluntarism on the part of the student in choosing his areas of study, the fact that there is some choice must certainly have something to do with the linkage between teaching and learning at the college level. Indeed it begins to explain the contemporary cry and demand for relevance that is abroad in the land today among college students critical of the course offerings and teaching methods in their institutions of higher learning. Especially under conditions of strong occupational motivation the sophisticated college student of this generation may demand courses of instruction which his

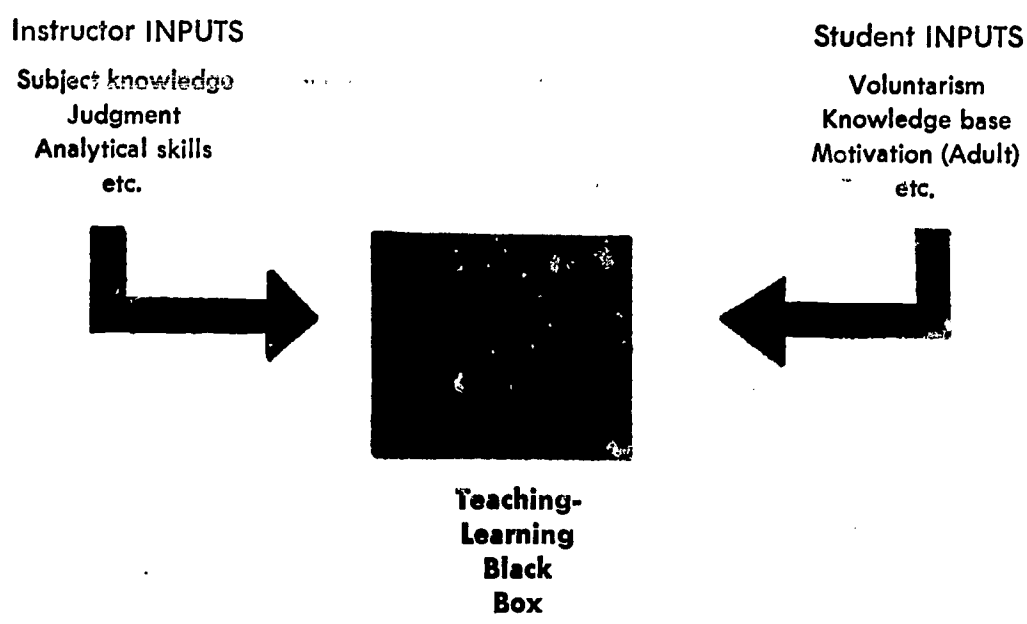
own professors are less than capable of offering. An underlying reason for the generation gap may be the fact that the students, who are supposed to be learning, may know different things, perhaps know them better, than the professors who are expected to do the teaching.

The choice behavior exercised by the student in selecting the subjects for his own instruction, and the knowledge base from which he is likely to both make this choice and evaluate the quality of instruction received, may be critical features of the student's participation in the teaching-learning process. At the college level we may have to model the teaching-learning link in a fashion quite different from the models of immediate post-natal stimulus-learning, or first grade teaching-learning, or even high school learning-teaching. Indeed, our own conclusions from this survey that independent study is as effective as any face-to-face instruction, when measured by student examination performance, should make clear that a great deal of learning resides in the student.

It is important to add a third dimension to the teaching-learning situation by examining the student contribution, and especially student motivation to learn. Much of the thinking about student motivation at pre-school and grade school levels does not seem to make much sense for college-age students. Those teaching techniques that derive directly from these preconceptions about student motivation at these early levels would, therefore, seem to be singularly inappropriate for college instruction. It is notable that the concepts of the "mature personality" capable of "self-realization" have built into them the notions of individual independence, self-discipline, and willingness to choose among goals and their means of attainment. Clearly, motivation to learn affects the balance of inputs into the teaching-learning black box among college-age students in the direction of increasing the amount of learning activity relative to the amount of teaching activity.

At least three factors characterize the teaching-learning situation at the college level: (1) voluntarism on the part of the student in choosing the subjects of instruction; (2) a knowledge base possessed by the student for making judgments about the content and quality of instruction received, judgments which, in turn, influence the voluntary choices made; and (3) the complex of culturally derived expectations and behaviors which comprise what we loosely summarize as the motivation to learn. Thus, a more accurate characterization of the input situation for the teaching-learning black box might look like the illustration below.

It is not our purpose in this state-of-the-art survey to develop a model of teaching-learning at the adult or near-adult levels. We do, however,



believe that anyone working or doing research at the college level of instruction can most readily make useful contributions if this linkage between teaching and learning becomes the center of their attention.

The evidence from this analysis conclusively demonstrates that differences among teaching methods occur only rarely as measured by a final examination. It seems safe to predict that any future studies replicating the four decades of research here summarized can only reproduce the results persented here. We have absolutely no reason to believe that tomorrow's studies, if they merely replicate what was done yesterday, will produce any results that differ to any significant degree.

The evidence is all in upon which we may base our conclusions about the relative utility of given methods of college teaching, *when this utility is measured through final examinations:* THERE ARE NO DIFFERENCES THAT AMOUNT TO ANYTHING.

The "Other Results" Conclusions

When repeated studies of comparative college teaching methods have revealed no differences between two or more methods, the researcher who anticipated results favoring his own preferred method is likely to allege that his preferred teaching technology produced "other results" that could be observed but not necessarily measured in the studies. Or, failing to observe these "other results," the researcher might piously conclude that had they been under observation the "other results" would have been measurable.

This is obviously a very constructive position to take for it assumes automatically that the measured outcome, student performance on final examinations, does not really measure or differentiate between the varied inputs of two or more teaching methods. Thus the claim for "other results" turns out to be a disguised form of plea for an adequate theory of teaching-learning in which the output variables are more imaginatively defined than by the final examination grade. The disappointing aspect of the research literature is that the promise of building a viable model of the teaching-learning process contained in the belief that "other results" were achieved by various preferred methods of instruction has never been realized.[1]

We are here involved in the fundamental relation between fact and ideology. A professor who believes that discussion or a tutorial is most productive of a learning consequence takes comfort in the fact that no research has *disproved* this conclusion, but he scarcely finds the support for his ideological position in any empirical data. The administrator who decides from a cost standpoint it is more efficient to use large lecture classes can also point to the research literature and say, "My position is not disproved." He cannot, however, just as his professorial antagonist cannot, claim that his conclusion is supported by the data. In this inconclusive situation the decision-maker's ideology will continue to dominate the operating conclusions by which programs of instruction at the college and university level are designed. We can only hope that bringing the state-of-the-art in college instruction up to date with respect to the research literature will make clear to those who choose among college teaching methods the ideological grounds for their choices. We also hope that it should now be clear that if research-based policy decisions are to be favored, that the research required to provide grounded policy decisions is yet to be done in the area of comparative college teaching methods.

Summary—The Problem Posed

The problem posed by this state-of-the-art review may be set forth simply.

---

[1] That this failing is not solely characteristic of educationists and educational researchers is tellingly revealed in the survey of the research literature made by Campbell and Dunnette in trying to tease out the measured consequences of sensitivity training. Their results show that none of the claimed advantages of sensitivity training had ever been measured in real situations and that either there were no particular advantages ascribable to sensitivity training or the outcomes measured were not relevant to these advantages. See: John P. Campbell and Marvin D. Dunnette, "The Effectiveness of T-Group Experiences in Managerial Training and Development," pre-publication copy. (The University of Minnesota, 1967, mimeographed)

Given a population of seven million adults or near-adults attending two- and four-year colleges and universities as students, what can we say about the relations among various methods for instructing them and the outcomes produced when measured on final examinations in their courses? Does the research literature give us a definitive answer to this question?

The results of our intensive reanalysis of data on comparative college teaching methods make it very clear that our intended goal has been achieved. We are able to state decisively that no particular method of college instruction is measurably to be preferred over another, when evaluated by student examination performances. We may also conclude that replication of the 91 studies examined in detail in this survey would not produce conclusions different from ours.

Any future research on comparative teaching methods at the college level must move in new directions. We have suggested that a fruitful direction of further analysis will be to examine directly the links between teaching and learning for a student group of adults or near-adults.

CHAPTER
2

The Sound and Fury, Signifying Nothing

Approximately four decades of research are represented in this analysis. The history of these research efforts began in the period just following World War I. There is concern throughout the four-decade period with using the methods of science to establish the superior utility of one college teaching method over another.

Two interesting themes are evident in this survey of the literature:

(1) The first and dominant theme is the optimistic assumption that a "scientific methodology" could be applied to the analysis of comparative teaching methods. It was assumed that by carefully employing experimental controls, the contrast between two or more teaching methods could be objectively set forth and the consequences measured, usually through examinations. The problem of experimental design was viewed as a very simple one—to clearly distinguish between two or more inputs and then to measure whether or not their respective outputs were alike or different. It did not matter whether the null assumption was made or whether a predicted difference in a given direction was set forth. In both instances the experimental design was identical.

(2) The second theme relates to seeking an explanation for the research findings. As the pile-up of results made clear that the preponder-

ance of evidence tested showed no differences, even between obviously distinctive teaching methods, the researchers turned attention to flaws in the research process to explain the non-conclusive results. If no differences resulted it was often alleged that this was because the measuring instruments were too gross to detect the "true" differences that "really" existed. Under similar circumstances, alternatively it might be argued that there was not quite enough difference between one teaching method and another as to produce the desired and expected differences in measured outcomes.

Throughout the four decades of research no one was willing to give up the belief that there had to be significant differences in outcomes produced by different college teaching methods. It has been only recently that the thoughtful researchers have begun to ask whether the fault lay with conceptualization of the problem rather than with the technologies of research employed. This, of course, is the issue we hope will be forced by the present monograph.

We think it is time for the researchers in the field to reconceptualize the problem as one of modeling the linkages between teaching and learning at the adult level. This becomes all the more pertinent because the present widespread use of educational TV, and the rapid development of programmed instructional methods demand a significant broadening of our models of learning to include unique teaching techniques. In television instruction the direct face-to-face relationship between instructor and student is eliminated. Yet, in both instances, there is still involved a teaching technology. In another monograph we review the extant studies of comparisons between college instruction by television and face-to-face methods. We conclude that there is no significant advantage to face-to-face instruction when measured against conventional instruction by television and we ascribe the differences which do appear to differences of *media* rather than differences in instructional *methods*.[1] We may also predict that as future comparative studies are made between programmed instruction and other teaching methods, the same outcomes will obtain.

The Persistence of Hope

The question must be raised as to why, in the face of mounting evidence, it was more than four decades before the initial analytical problem was reformulated. Why didn't the accumulated research speak to each new generation so that it would not replicate the same inconclusive studies of its ancestors?

---

[1] See: Robert Dubin and R. Alan Hedley, *The Medium May Be Related to the Message: College Instruction by TV* (Eugene, Oregon: CASEA, 1969).

To this question one might simply answer, "When the facts get in the way of an ideological conviction, the facts must be bent or discarded, but the conviction should remain unchanged." Thus, if college professors are convinced that tutorials or discussions are most productive of learning, and the research data do not disprove that conviction, it seems fair to continue to hold to that belief. The fact that the research findings do not prove that tutorials and discussions are superior is taken as irrelevant information. It leads to the conclusion: "If I am not proved wrong, then I have every reason to believe that I am still right." But no one else was proved wrong either in their espousal of other teaching methods. The net effect of the inconclusive data was to permit each person to hold to his private preference concerning a given teaching method without the data demanding its alteration.

There must be other reasons for the dominance of ideological justifications for particular teaching methods right up to the present time. We will examine three influences: (1) the persistence of traditions about teaching; (2) the pressures for changes in teaching methods generated by rapidly growing enrollment and slowly increasing teaching faculties; and (3) the influences of ideological developments largely external to the institution of higher education.

The traditions of a profession, like college teaching, run deeply and change slowly. The generation of professors who were trained and teaching before and right up through World War II had largely experienced as students, and later in their teaching careers, small-class, tutorial-like, teaching situations. Small classes were central to the basic tradition of colleges and even universities. Where large lectures were employed they provided the exception in teaching practice in order that many students might enjoy the benefits of a brilliant lecturer or renowned figure. It seemed justified to conserve the time or the energies of the distinguished by permitting their performance in the lecture amphitheater. The situation changed after World War II, but the very need to move to larger classes and more impersonal relations between faculty and students in the classroom could well have confirmed, precisely because it seemed to be an exception to the tradition, the continuing preference for the historically certified "right" way to teach college students. Therefore, we have reason to believe that right up to the present time the modal preference of college professors is for the small size class, discussion-type interaction with students in a personalistic fashion.

A significant structural influence on the preference for particular kinds of teaching methods was the rate of change in the growth of faculties in

relation to student bodies. Prior to the end of World War II the supply of professors grew at a rate comparable to the rate of growth of the student body. Under these circumstances the student-teacher ratio remained largely unchanged. From 1940 to 1945 the withdrawal of young men from the campus to serve in the armed forces actually reduced the student-faculty ratio.

In the post-war era the pressure of very rapidly increasing rate of student enrollment was not matched by the growth of teaching faculty. Not only did an increasing proportion of professionally trained faculty members devote part or full time to research, but even those who were teaching exclusively experienced a decline in hours-per-week teaching load. Combined, these factors forced enlargement of classes and the utilization of teaching methods such as closed circuit television. The mass methods of instruction were invariably viewed as compromising "true" educational experiences in the interest of expediency in order to handle the increasing enrollments. Thus, teaching more students without a comparable increase in instructors necessitated a move away from the ideal of small classes with a Mr. Chips at the head of the seminar table. The change was viewed as a necessary evil and therefore no genuine legitimation of the new teaching methods was attempted or achieved. It is not surprising therefore that preferences for the "classical" methods of college instruction should persist even in the face of demonstrated need for inventing or considering alternatives.

Developments that took place essentially outside the academy but had repercussions upon it comprised a third influence upon the ideological commitments of American university professors to classical methods of college instruction. The post-World War I period was one in which psychologists gave special emphasis to individual differences and the need for designing educational opportunities to emphasize and realize the differing potentials of individuals. The progressive education movement had its roots in an empirical philosophy of the individual. It seems obvious that the informal, student-centered instructional methods would be most consistent with a belief in the desirability of maximizing the potential of individuals of differing talents and abilities. We may then conclude that in the period between the world wars, the philosophic orientation of progressive education, grounded in the psychology of individual differences, would have its repercussions upon the colleges and universities, at least with respect to teaching methods or preference for teaching methods.

In the 1950's a similar philosophic movement reached it peak, a movement characterized by a commitment to what came to be called group

dynamics and by a belief that the individual, being the product of group experiences, achieved his maximum self-realization in small group contexts. Indeed, this philosophic movement remains in current high fashion, under the name of sensitivity training. In this philosophic tradition the pedagogy most adaptable to maximizing small group experiences is the small group itself, with the instructor operating as a resource to the experiencing group.

These successive philosophic traditions, although relatively different from each other, both had identical consequences for ideological preferences of university professors for a given method of teaching.

By combining these three strands of analysis we may conclude that the traditional beliefs about teaching, the relations between demand for teaching services and their supply, and the impact of philosophic positions upon professors' views of the teaching-learning linkage, all served to reinforce the generally-held belief that discussion and tutorial methods of instruction were to be preferred at the university and college level.[2]

The ideological conviction that tutorial and small group situations were most efficacious for college-level teaching led to studies and experiments to demonstrate that this was, in fact, a scientifically grounded conclusion. The motive for doing the comparative studies was to prove a point already well established in the minds of professors. Only when demands could not be satisfied within traditional methods were professors willing to move to other methods of instruction, including large lecture classes, educational TV, and ultimately, programmed instruction. There was no experimental evidence that supported the utility of these alternate methods over the traditional modes of college instruction. Nor was there, and this should be emphasized, any evidence that these alternate modes of instruction were any worse than the traditional ones.

The knowledge base for making innovations in instructional methods was neutral with respect to the particular innovations considered. Perhaps this very neutrality of the knowledge for making an informed decision based upon the measured results of given methods of instruction is the key to understanding how, in the face of a very strong and determined ideological conviction to the contrary, large classes, impersonal instruc-

[2] It should be strongly emphasized here that since the data from which we are to reach our conclusions are based on undergraduate instruction in colleges and universities, we have not drawn into the argument any conclusions based upon the principal mode of instruction at the graduate level. This, however, would also have an impact in exactly the same direction as the three factors we have already examined because much of graduate instruction approximates an apprenticeship in which the individual instructor and the individual student have more or less continuous interaction.

tional relationships, and even machine teaching, were able to gain a significant place in the repertory of college teaching methods. Put another way, the rate of social change may be markedly increased by the very lack of evidence that a given system of behavior is either better or worse than an alternative.

Ideological Roots of Comparative Research

We here simply present a straightforward recital of what the various researchers did, and in many cases why they did it, in the course of making comparative analyses of college teaching methods. The account is chronological so that the reader may have a sense of the history of ideas that have characterized the justification for comparative research.

We have already pointed out that the progressive education movement influenced ideas about college teaching. We might therefore expect that some of the early studies of comparative college teaching methods would come from courses in education and/or psychology.

In 1924 we had the following study:

> During the first semester of the school year 1922-23 an experiment was conducted by the authors to determine what differences existed in the measurable results of instruction in a large (lecture) and small (discussion) class in an education course offered in the School of Education of the University of Michigan. (p. 1) . . . results in terms of semester averages indicate no appreciable difference in the achievement of the large and small class included in this study. (p. 12)[3]

This is one of the early studies to find that no differences indeed existed.

The influence of Goodwin Watson at Teachers College is clearly revealed in the next report.

> The present experiment is one of the attempts which Dr. Goodwin B. Watson and the author have made to secure better results from the teaching of Educational Psychology. The lecture method did not seem to recommend itself as wholly desirable and yet, as it is such a common type of teaching it certainly was desirable to compare other plans with it. The experiment, therefore, set out to compare the lecture method with a plan involving class discussion. (p. 454)
> ... we can only conclude that for two large (N=120-170) sections of graduate students in Educational Psychology, meeting once a week for two hours, with little previous experience or feeling in favor of the discussion method, the lecture method is superior to the discussion method in producing improvement in things measured by tests. (p. 461)
> The problem that must be solved is not the question, "Is Method A

[3] J. B. Edmundson and F. J. Mulder, "Size of class as a factor in university instruction," *Journal of Educational Research*, 9 (1924), pp. 1-12.

> better than Method B or Method C?" but rather, "What are the conditions under which Method A produces most effective results? What are the situations where Method B is best?, etc." There are times when the lecture method is an effective aide in securing desirable changes in students. There are other times when the lecture method is harmful and should be replaced by discussion or individual projects or something else. (p. 462)[4]

But this time it had already become apparent that the analytical problem was not the old industrial management problem a la Taylor of discovering the "one best method" but rather asking the question, does each method employed have its unique advantages? It is notable that this question persists through the literature and always remains unanswered in terms of empirical data. Spence had no data on which to base his final conclusion, nor do succeeding authors who insist that each method has its unique advantages.

From this point on there is constant surprise expressed because the expected differences in favor of traditional teaching methods do not emerge. The following study reports an analysis of the teaching of English History:

> This study was planned and carried out with a view to throwing some light on the question of the value of quiz sections (discussion methods) in the teaching of history. Two sections of English History were given during the year of 1927-28, one meeting three times a week for lectures only, the other meeting three times a week for lectures and once a week in small groups. (p. 276)
> . . . we are led to a conclusion contrary to all expectations—namely, that the added hour a week devoted to quiz sections had no significant value. (p. 282)[5]

In the depths of the depression the Department of Psychology at the University of Minnesota determined to investigate the relative utility of lecture-quiz and all-lecture methods of instruction. For the introductory course the following results obtained:

> In view of the available experimental evidence and the increasing problems brought about by quiz sections, the Department of Psychology decided to conduct a controlled investigation to ascertain, if possible, the relative merits of the lecture-quiz and all-lecture method. (p. 33)
> From the . . . findings we may conclude that under the conditions sur-

[4] R. B. Spence, "Lecture and class discussion in teaching educational psychology," *Journal of Educational Pyschology*, 19 (1928), pp. 454-462.

[5] D. G. Barnes, and H. R. Douglass, "The value of extra quiz sections in the teaching of history," *U. of Oregon Publications, Educational Series I: 1929*, 1, No. 1, pp. 276-284.

rounding this experiment, there is no difference in the value of the two methods employed. (p. 48)[6]

In this study Longstaff reviewed the literature up to 1932 and in 1942 and in 1955 the distinguished psychologists Dale Wolfle and Birney and McKeachie reviewed the research literature on comparative college teaching methods and concluded that Longstaff's 1932 review still accurately described the results of research. As Birney and McKeachie say:

> *Teaching methods: summary.* In 1942, Wolfle summarized research up to that time by repeating Longstaff's statement of 1932: "The experimental evidence submitted to the present time tends to support the general conclusion that there is little difference in achievement in large and small classes and, also, that it makes little difference as to what method of presentation of the materials of the course is used." [D. L. Wolfle, "The first course in psychology," *Psychological Bulletin*, 39, (1942), pp. 685-712] The third decade of research has not outdated Longstaff's statement. However, recent research does hold forth the promise that in the next decade we will have a better understanding of the effect of various teaching methods on student learning.[7]

Thus we have in neat historical order three decades of research reviewed with the clear-cut conclusion that no demonstrated differences can be found in the research literature with respect to college teaching methods. But the final reviewer in 1955 still maintains an optimistic expectation that such differences eventually will emerge from the future research.

By 1940 the research results were being clearly incorporated into the textbooks on college teaching. For example, Cole reached the following conclusion after summarizing ten studies (there being many more already in the field):

> In so far as mere mastery of elementary subject matter is concerned, the lecture method is as good as any other. The results are sometimes almost identical for the lectures and discussion groups, sometimes slightly in favor of the discussion classes, and often appreciably in favor of the lecture classes.[8]

In the post-World War II period we can see emerging the influence of the group dynamics movement in giving direction to the formulation of research problems. Thus, for example, the term non-directive tends to be

---

[6] H. P. Longstaff, "Analysis of some factors conditioning learning in general psychology," *Journal of Applied Psychology*, 16, (1932), pp. 9-48; 131-166.

[7] R. Birney, and W. McKeachie, "The teaching of psychology: A survey of research since 1942," *Psychological Bulletin*, 52, No. 1, (1955) p. 58.

[8] L. Cole, *The Background for College Teaching* (New York: Farrar & Rinehart, 1940), pp. 324-325.

used in place of "discussion method" or "tutorial." The following is a typical description of the new direction for research on college teaching.

> The aim of this study was to evaluate the over-all effectiveness of non-directive teaching of an undergraduate course in general psychology as compared to the traditional lecture-discussion. (p. 19)
> The results of this phase of the experiment indicate that non-directive teaching, as outlined in this study, is not as effective as the traditional teaching techniques in helping students master the factual subject matter of a course in general psychology as measured by an examination based on knowledge of textbook material.[9]

Given the reality of large lecture classes, especially in introductory courses, studies emerge in which, perhaps to salve the conscience, it is attempted to find out whether the continuing belief in small intimate classes is justified. Thus we have the following report on the introductory course in psychology.

> The present investigation was originally undertaken as a local report to the department and the administration to see how well or how poorly students exposed . . . to impersonal, mass, education might do. (p. 298)
> The main results were as follows: The lecture sections averaged for the six quarters three points higher than the small classes, with superiority appearing five quarters out of six, and the sixth coming out a tie. In no case did the smaller group earn a higher course total. (p. 299)
> In spite of all . . . arguments in favor of large lecture sections, which are largely based on statistics on grade-point earnings, the writer must admit that he is not 100 per cent convinced of the efficacy of such "mass education." (p. 300)[10]

What is, of course, interesting is that the writer refuses to be "100 per cent convinced of the efficacy of such 'mass education.'" The point is not that the superiority of the large class has been demonstrated since the differences are not statistically significant as reported in this study. What the writer is still not willing to admit is that the presumed superiority of the small discussion class is also not proved by this or preceding studies.

The following is an illustration of how the language changes but the problem remains the same.

> This study limited itself to the investigation of two styles of teaching, operationally defined below as "directive" and "permissive."
> *Summary and Conclusions*— The D and P groups showed no significant differences on the objective-final when taken as a whole, but when the

[9] M. J. Asch, "Non-directive teaching in psychology: An experimental study," *Psychological Monographs*, 45, No. 4, (1951), pp. 1-24.

[10] Richard W. Husband, "A statistical comparison of the efficacy of large lecture versus smaller recitation sections upon achievement in general psychology," *Journal of Psychology*, 31, (1951), pp. 297-300.

> two teaching methods were analyzed for their effects on the "better students" and the "poorer students" it was demonstrated that the directive sections were more beneficial to the poorer student. (p. 184)[11]

In Wispe's study the comparison is now between "directive" and "permissive" teaching styles. The conclusion reached is the standard one of "no difference." But there is a suggestion that when student bodies are differentiated according to their ability, the "directive" method may be superior.

By 1954, Guetzkow, Kelly, and McKeachie experimentally studied three distinctive methods of college teaching. Their description of the research and their conclusions are self-explanatory.

> ...the design may be stated in the null form: There will be no significant differences in outcome with respect to course objectives between students whose class meetings are taught by Method 1 (recitation-drill) and those students who are taught by Method 2 (group-discussion) or by Method 3 (tutorials). (p. 195)
>
> By and large, we found no differences between the three teaching methods from the point of view of educational outcome. The few statistical differences, in general, favored the recitation-drill method. The results of the experiment, of course, were contrary to our original expectations. They constitute a clear confirmation of the general conclusion derived from experiments on instructional procedures since the early 20's. As Good recently put it, after reviewing the literature in this area, "The complexity of the teaching-learning process is such that attempts to establish the relative merit of a 'general method' of teaching are likely to prove inconclusive." [C. V. Good, "Colleges and universities: VIII. Methods of teaching," in W. S. Monroe (Ed.), *Encyclopedia of Educational Research*, New York: Macmillan, 1952, pp. 273-278.] (pp. 205-206)[12]

Thirty years after the first study cited here we are still getting the same conclusions. Guetzkow, Kelly and McKeachie at least maintained the posture of optimism that must surely sustain researchers after thirty years of failure to demonstrate what they expect to find.

As if to recognize that there is indeed an ideological commitment to given methods of instruction and irrespective of the research findings, McKeachie followed up the work he did with Guetzkow and Kelly in the same year and reached the following conclusions.

> One would expect that the controversy between our education's authoritarians and permissivists would long ago have been resolved by the

[11] L. G. Wispe, "Evaluating section teaching methods in the introductory course," *Journal of Educational Research*, 45, (1951), pp. 164-186.

[12] H. Guetzkow, L. E. Kelly, and W. J. McKeachie, "An experimental comparison of recitation, discussion, and tutorial methods in college teaching," *Journal of Educational Psychology*, 45, (1954) pp. 193-207.

cold logic of experimental studies. Unfortunately, this just hasn't happened. The published experimental studies are not in agreement and there are a host of unpublished studies which remain unpublished because the two methods used produced no significant differences in outcomes. (p. 146)

What are we to conclude from these studies? While there is a dearth of follow-up data, with such slender results at the end of the courses our hope that either method produces significantly greater long-time benefits is probably unrealistic. Should everyone go his own way and teach any way he pleases? Personally, we are not willing to go quite so far, but certainly none of us should exclaim with horror, "His classes are instructor-centered." As psychologists, however, we believe in research. Why has research on student-centered versus instructor-centered teaching seemed to lead up a blind alley? (p. 148)[13]

In a 1958 review of 66 research reports the Bureau of Institutional Research at the University of Minnesota reached the following conclusion.

Undoubtedly the most striking finding of this review is the consistent inability of investigators to demonstrate statistically significant differences between the experimental and control methods of teaching. This seemed to hold true regardless of the subject field or class size being examined. In the few studies that did suggest significant findings the differences were generally small and of doubtful practical significance. Unfortunately the finding of no significant differences does not legitimately allow one to conclude that differences in teaching efficacy are not related to method and/or class size. Rather, the fact that essentially none of the investigations have been able to demonstrate practical differences suggests that some careful examination of the experimental methodology and evaluation procedures may be in order.[14]

It will be noted here that a direction is being suggested in which the methodology of research rather than the conceptualization of the research problem is considered the most fruitful line of further investigation.

The preoccupation with methodology is suggested again in the same year by Nachman and Opochinsky.

Reviews of teaching research have consistently concluded that different teaching procedures produce little or no difference in the amount of knowledge gained by the students. This same conclusion has been reached despite the fact that experimenters have employed a wide variety of independent variables, such as lecture versus discussion classes, instructor-centered vs. student-centered classes, large versus small classes, var-

[13] W. J. McKeachie, "Student-centered versus instructor-centered instruction," *Journal of Educational Psychology*, 45, (1954). McKeachie went on to suggest some of the reasons why the research results have been inconclusive.

[14] Bureau of Institutional Research, "A Review of the Literature Concerning Studies of College Teaching Methods and Class size," University of Minnesota, 1958, p. 4. (mimeographed)

> ious types of TV classes, etc. These results are surprising if one considers that much of the research was instigated by the hypothesis that differences would be found. Furthermore, it appears as if most educators still assume that classroom techniques do in fact have specific effects. Why then have differences not been found?
> The purpose of this paper is to examine an alternative hypothesis, namely, that the different teaching methods have, in fact, produced differential amounts of learning but that these effects have been masked in the measurement process.[15]

We started in 1924 and we end with a quotation from the widely used *Handbook of Research on Teaching* published in 1963, almost 40 years after the initial quotation.

> The effectiveness of lecture and discussion methods has often been compared. Since discussion offers the opportunity for a good deal of student activity and feedback, it could, in theory, be more effective than the lecture method in developing concepts and problem-solving skills. However, since the rate of transmission of information is slow in discussion classes, we would expect lecture classes to be superior in helping students acquire knowledge. (p. 1126)
> Despite the many findings of no significant differences in effectiveness between lecture and discussion, those studies which have found differences make surprisingly good sense. In only two studies was one method superior to the other on a measure of knowledge of subject matter; both studies favored the lecture method. In all six experiments finding significant differences favoring discussion over lecture, the measures were other than final examinations testing knowledge . . .
> When one is asked whether lecture is better than discussion, the appropriate counter would seem to be, "For what goals?"[16]

We have here, of course, reviewed many more studies than those covered by Professor McKeachie in his analysis in the *Handbook*. Professor McKeachie may be quite right in suggesting that the rare cases where differences favor one method of teaching over another provide us with possible clues as to why. It is clear, however, that McKeachie prefers the discussion method, based upon the work of researchers, like N. R. F. Maier, who have focused their analytical attention solely on discussion groups, rather than being grounded in comparative studies of two or more teaching methods.[17]

---

[15] M. Nachman, and S. Opochinsky, "The effects of different teaching methods: A methodological study," *Journal of Educational Psychology*, 49, No. 5, (1958), pp. 245-249.

[16] W. J. McKeachie, "Research on teaching at the college and university level," in N. L. Gage (Ed.), *Handbook of Research on Teaching* (Chicago: Rand McNally & Co., 1963) pp. 1126-1127.

[17] See W. J. McKeachie, "The discussion group," *Memo to the Faculty*, No. 14. (Ann Arbor, Michigan: Center for Research on Learning and Teaching, 1965).

The purpose of this brief review is not to prove how stupid, or narrow-minded, or ideologically committed have been the researchers who have compared college teaching methods. We simply want to demonstrate that the data are overwhelming in the direction of no differences among various methods of college instruction. In the next chapter we make one last massive effort to demonstrate this all over again. We hope our demonstration, because it reanalyzes the data rather than the conclusions of the various studies, will be all that much more convincing than have previous summaries. We are convinced that approximately 40 years of research speaks the truth. It is now time to turn to a reconceptualization of the analytical problem.

We can no longer be satisfied that there are pedagogical theories that confirm and predict the advantage of one teaching method over another. We are now convinced that the proper conceptualization of the problem, already foreshadowed in the Guetzkow, Kelly, McKeachie summary, is to build a model or models of the learning-teaching processes in which pedagogy is only one input into the process, although admittedly a complex one.

CHAPTER
3

There Are No Differences

Academic policies regarding the use of various methods of teaching at the college level should be grounded in research. The data exist upon which to base grounded policy decisions. In this chapter we present the summary conclusions setting forth the results of comparative analyses of college teaching methods. The methods utilized in obtaining these results are presented in APPENDIX A, and the detailed results not reported in this chapter are contained in APPENDIX B.

How It Was Done

All previous summaries of the results of individual comparative studies added up the *conclusions* of separate researchers. The output of such summaries was, therefore, a set of conclusions about conclusions.

We did not follow this procedure. Our analysis is based upon the actual data contained in the individual comparative studies. It is our belief that sometimes researchers mistake their own data and reach faulty conclusions. We wished to avoid this source of error.

The data utilized fell into two categories. In a number of studies (18), the author only reported the mean performance scores of groups taught by different methods. These data could only be used to determine that one

method of instruction produced better examination results than another, or that there were no differences measured. These studies, therefore, reported only the sign of the differences. The magnitude of the differences was unimportant. Such data are the least precise, but provide us with the opportunity to make a *sign test* (over many studies, is one method of instruction favored over another regardless of the amount of measured difference in examination scores?).

Where data on the distribution of scores (means, standard deviations, sample sizes) for the groups compared were available, we were able to use or calculate a standardized statistical measure of the differences in *all* possible mean examination scores. This measure provided high precision in determining the actual amount of difference between groups' scores in each comparison.

This identical procedure was then repeated for all the studies where sufficient data were available, employing only the *independent* comparisons. In many studies a number of comparisons were made in which the same groups were compared several times. Thus, results on mid-term examinations and final examinations might be reported. We limited independent comparisons to a single comparison, usually the final examination results, between two groups of students. Thus, even though comparisons between the same two groups were reported in a study for midterm and final examinations or for term papers and examinations, we chose only the final examination results as being independent. We thus eliminated repeated comparisons of the same two student groups which obviously are not independent.

In this chapter each chart and the accompanying statistics report the results of the independent comparisons. Each chart shows the actual distribution of standardized differences and reports the statistics for the distribution as well as the sign test for the independent comparisons. The footnote to each chart reports the sign tests for all differences measured.

We thus have four different analyses of the data.

(1) The sign test of *all* differences measured, representing the greatest number of possible comparisons.

(2) The distribution of standardized differences in *all* measured group examination scores and associated statistics, which had the next largest number of possible comparisons.

(3) The sign test of those comparisons in all of the studies which were considered *independent* comparisons, with a smaller number of cases than the previous two categories.

(4) The distribution of standardized differences in measured group examination scores and associated statistics for *independent* comparisons only, the smallest number of comparisons utilized.

Briefly then, our method was to use examination data employing objective measures of the differences obtained in group examination scores. We proceeded from the least precise objective measure to the most precise objective measure, ending up with four distinctive ways to measure the differences in student performance for every comparison analyzed.

The entire analysis involved reading every study carefully, abstracting the data from it, translating the data to a form capable of computer processing, and designing computer programs to analyze the results.[1]

It Pays to Study

The most obvious comparison that may be made in determining whether anything makes a difference in examination performance by college students is to contrast a group of students who studied with those who did not. Throughout the entire period of his schooling the student is reminded of the desirability of studying. The terms "school" and "study" almost become synonymous. Belief in this identity is often shared by schoolmen. This belief is so strong that few functionaries of schools, or researchers on them, have ever done research to confirm or challenge the identity.

We found two studies in the literature which compared some form of study with no study and evaluated their respective outcomes on examinations covering ability to recall or prove knowledge of course content. These studies had a total of six comparisons between groups of students who studied and those who did not, all of which were independent comparisons. The results are significantly in favor of study.[2]

When it comes to learning a specific subject matter we have some modest assurance that for college students studying *does* make a difference. This should be a very comforting conclusion for those who believe that a college or university education, and study, are somehow or other linked. Whatever there is about the college and university environment that makes studying by students a more probable activity than not studying, must be taken as an important contribution of the higher educational

[1] The computer programs utilized and the detailed results are available to any user by application to the librarian, Center for the Advanced Study of Educational Administration, University of Oregon, Eugene, Oregon 97403. Data will be provided on a cost basis.

[2] Studies number 057 and 062 report in every comparison that studying produces far superior examination grades than could be achieved by students who did not study the course content. The studies are listed by number in the bibliography.

institution. Perhaps, this may even be its most important contribution—to develop the habits of study which are, or may be, the precondition of learning. At least we can start with meager data which are definitive in establishing the value of study when compared with no study in influencing college student grades.

Teaching Method Comparisons

We have organized our comparisons among teaching methods along a scale which at one end emphasizes instructor-centered classroom situations, and at the other end emphasizes student-centered learning situations which may or may not be found in a classroom setting. Thus, we proceed systematically from an examination of lecture only compared to other instructional methods, to a comparison of independent study with those methods with which it has been contrasted. It will perhaps help to orient the reader if we briefly describe the characteristics represented by this broad range of teaching-learning situations.

To reiterate a point made in the first chapter, we believe that teaching is a technology. Teaching consists of a set of ideas to be communicated, models of the teaching process and implicit assumptions concerning the teaching-learning linkage, and a set of instructional activities summed up as the "method" of teaching.

Researchers have been attempting to compare and contrast different college-level teaching technologies for approximately 40 years. We may classify these technologies along a number of different, yet overlapping, dimensions. For example, two dimensions are the extent to which emphasis is placed upon the instructional activities of the teacher, or the degree to which learning is considered to be a social event for which group behavior is important.

In this chapter, we present the results of experimental comparisons of different teaching technologies. In so doing, we shall briefly discuss some of the characteristic features of the different technologies which have received attention in the research literature.

FACE-TO-FACE INSTRUCTION

Face-to-face methods of teaching such as the lecture, group-discussion, and the tutorial are explicit examples of teaching technologies based on traditional assumptions concerning the teaching-learning linkage. The instructor is not only assumed to be a sufficient condition but also a necessary condition for learning, and outputs from the teaching-learning situation are assumed to be a function of differential teaching inputs.

The lecture or "telling" method is the most widely used method of teaching in colleges and universities today. The popularity of this method probably derives from the need to handle increasing enrollments in institutions of higher learning, although it is also tied to the belief that the fundamental goal of an instructor is to transmit knowledge.[3]

The lecture method assumes the superior knowledge of the lecturer and therefore places in his hands the selection of subject matter and the depth of coverage of the subject matter, the balance between content and illustration, the length of the lecture period, and the frequency of the lectures in a given time period.

Lecture classes are usually,

> . . . dominated by an extended discourse by the teacher but . . . [they] . . . typically include, in addition, assignment of readings and measurement of achievement. [They] . . . may also include oral quizzes or recitations. The lecture may be informal, affording opportunity for students to ask questions. The teacher may ask a few questions to stimulate interest and focus attention on the topic to be dealt with.[4]

The principal emphasis in lecture classes is on the instructional activities of the teacher, not on the activities of the students.

A number of other face-to-face methods of teaching have developed in reaction to the "sterile," "authoritarian" image of the traditional lecture. To a greater or lesser degree, these methods add to the assumption that the lecturer possesses superior knowledge other assumptions: (1) that students should be active rather than passive participants in the teaching-learning situation; (2) that opportunities should be provided for the students to clear up hazy points, and correct faulty learning; and (3) that attempts should be made to maximize student motivation and interest. Learning is considered to be, ". . . an event of social interaction," an outcome of "interpersonal-encounter" between teacher and students. Thus, emphasis in the classroom is shifted from the instructional activities of the teacher to the interaction between teacher and students and to procedures for "socializing" the teaching-learning situation.

The three most commonly used methods which derive from this model of the teaching-learning process are the discussion method, certain combinations of the lecture and discussion methods, and the tutorial method. As the name implies, discussion classes primarily involve an interchange

---

[3] W. J. McKeachie, "Research on teaching at the college and university level," in N. L. Gage (Ed.), *Handbook of Research on Teaching.* (Chicago: Rand McNally & Co., 1963), pp. 1125-1126.

[4] W. Monroe, and A. Marks, "General methods of teaching," *Educational Administration and Supervision,* 24 (1938), p. 498.

of ideas and knowledge between a teacher and a group of students. Although there are a number of variations on the discussion theme, the overriding goal is maximization of self-realization on the part of students through participation in small-group contexts.

A common practice in many colleges and universities today is to, ". . . divide class time between lectures and discussions."[5] The lecture sessions are used to transmit factual information rapidly and efficiently, while the discussion sessions provide opportunity for student activity and feedback, and for encouraging student interest and motivation. This is the "have-your-cake-and-eat-it-too method."

If asked what they considered to be the ideal method of teaching, many college teachers would probably specify the tutorial method. The tutorial involves regular interpersonal conferences between a student and a teacher. During these conferences discussion is focused on a reading or group of readings which the student has recently completed, or the teacher might engage in short discourses to guide the student in his attempt to master some particular aspect of the subject matter. Although this method may have virtue, it is sparingly used, except for graduate reading seminars, since it is extremely demanding of an instructor's time and energy and a costly means of instruction.

INDEPENDENT STUDY

In contrast to methods of face-to-face teaching, methods of independent study shift the focus of attention from the teacher and classroom interaction to the student as self-teacher. The principal assumption underlying these technologies is that, ". . . learning is an individual act, a set of events which take place entirely within the learner."[6]

> By definition, independent study delegates to the student primary responsibility for his own learning, and it is found that in practice very few restrictions are employed.[7]

The significant point here is that stress is shifted either partially or completely away from the activities of teaching by teachers to the activities of the student teaching himself and learning.

There are two major types of independent study which have received attention in the research literature. The first retains the idea of the neces-

---

[5] McKeachie, *op. cit.*, p. 1127.

[6] R. M. Gagne, "Learning research and its implications for independent learning," in G. T. Gleason (Ed.), *The Theory and Nature of Independent Learning* (Scranton, Pennsylvania: International Textbook Co., 1967), p. 30.

[7] B. D. Felder, "Characteristics of independent study practices in colleges and universities," (Doctoral Dissertation, University of Texas, 1963, Abstract).

sity for teacher-direction and guidance in the learning process, the second puts emphasis solely on the learner and upon the fact that, ". . . learning can, and often does take place in the absence of the teacher."[8]

Most schools have institutionalized independent study practices as *supervised* independent study in which ". . . students . . . receive regular guidance and supervision by faculty members."[9] This is typically the case when, for example, students use an independent study device such as a programmed booklet or an "autotutor" in class. The "individual-laboratory" is another form of supervised independent study in which the student works at his own pace but is required to attend class meetings.

The responsibilities of the teacher in an independent study class involve,

> . . . presenting the stimulus, . . . directing attention, . . . communicating a model of expected performance, . . . providing what we have called "learning guidance," . . . promoting transfer of learning, and . . . assessing the outcomes of learning. All of these functions are essential to the activity called instruction. Besides these, still other functions of tremendous importance to education are performed by the teacher, including particularly the enhancement of motivation, and the imparting of values.[10]

Thus, the emphasis is on the learner, but the teacher is retained as a motivator, a transmitter of values, and a guide to outlines, indexes, reference lists and other materials or devices designed for imparting knowledge to the student.

A final method we shall examine is *unsupervised* independent study which places sole responsibility on the student. Interaction here is between student and printed or other instructional materials (e.g. computer-assisted instruction), and with other students outside of the traditional classroom. Typically, at the outset of a course, students exposed to this method of instruction are given relevant bibliographies, and other learning aids and they then must fare for themselves. The role of the instructor is limited to constructing bibliographies and materials relevant to the particular subject matter.

There are two basic types of unsupervised independent study which have received attention in the research literature. The most common type, which we have arbitrarily designated "self-study," involves complete independence on the part of each student in working through subject matter. A second type of unsupervised independent study labeled "autonomous small groups" involves a group of students studying the subject matter together without benefit of classroom participation or of teacher-supervision.

---

[8] Gagne, *op. cit.*, p. 30.
[9] Felder, *op. cit.*, Abstract.
[10] Gagne, *op. cit.*, p. 30.

If There Is No Difference, It Makes No Difference

For approximately four decades researchers have been attempting to compare and contrast face-to-face methods of teaching, experimentally, in college and university classrooms. The central problem for research has been defined as one of finding the method or methods of teaching which will maximize desired consequences among students. The most widely used measure of output from the teaching-learning situation has been a final examination covering course-related subject matter. The results of this research are clear and unequivocal—no particular method of teaching is measurably to be preferred over another when evaluated by student examination performances.

VARIETIES OF FACE-TO-FACE INSTRUCTION

The most popular comparison which has been made repeatedly throughout the 40 year period is between the lecture and discussion methods of teaching. We have examined 45 studies which report comparisons of the two methods. Thirty-six of these studies present data which we have used in summarizing the results for this type of comparison. (See FIGURE 1, page 36.)

Of a total of 88 independent comparisons of the lecture and discussion method of teaching reported in the 36 studies reviewed, 45 (51.1%) favored the lecture method and 43 (48.9%) favored the discussion method. The distribution of standardized differences (actual differences in group mean scores divided by the estimate of the standard error of differences) between these methods is shown in FIGURE 1. The average difference of standardized scores is 0.09 and the standard deviation of differences is 1.70, which is not significantly different from a "true" difference of zero ($t = 0.38$, $P > .50$). Thus, we feel confident in concluding that the lecture and discussion are equally effective methods of instruction.[11]

The same conclusion holds for comparisons of the lecture method with combinations of the lecture and discussion method. Here we have examined 7 studies which present comparative data.[12] Of a total of 8 independent comparisons, 3 (37.5%) favor the lecture method, 4 (50.0%) favor the combination method, and 1 (12.5%) shows no difference in the achievement of students taught by the two methods. Although there

[11] The technical details of the measures used to determine whether there are any significant differences between teaching methods compared are discussed fully in APPENDIX A.

[12] None of the comparisons reported here dealing with discussion methods was included in the preceding analysis of lecture *vs.* discussion methods.

appears to be a real difference in the number of comparisons favoring each method, the average difference is only 0.19 with a standard deviation of 0.84. Here again, the average difference is not significantly different from a "true" difference of zero ($t=0.46$, $P>.50$). The distribution of standardized differences is presented in FIGURE 2. (See page 37.)

Three studies were examined which compared the discussion method with combinations of the lecture and discussion methods. Of 12 independent comparisons reported in these studies, 5 (41.7%) favored the discussion method, 5 (41.7%) favored the combination method and 2 (16.6%) showed no difference in mean achievement. The mean of the standardized differences is 0.20 and the standard deviation of the differences is 1.53. This mean difference is not significantly distinct from a "true" mean difference of zero ($t=0.43$, $P>.50$). The distribution of standardized differences for these comparisons is shown in FIGURE 3. (See page 38.)

We omit a discussion of comparisons involving different combinations of the lecture and discussion methods. The data for these separate types of comparisons are presented in APPENDIX B. These results uniformly support the no-difference conclusion.

Although there has not been a large number of experimental comparisons involving the "ideal method," the tutorial, we have examined data from 3 independent comparisons of the tutorial with the lecture method and from 2 independent comparisons of the tutorial with the discussion method. The data, which are presented in APPENDIX B are clear in suggesting that none of three standardized differences is significantly different from a "true" difference of zero.

In sum, we have reviewed the data from experimental studies which were attempts to compare a number of different types of face-to-face methods of teaching. In every instance, the accumulated data from these studies indicated that the methods of teaching compared were not measurably different when evaluated by student examination performances. These data suggest that many of the time-worn theories of the teaching-learning linkage are simply inadequate. We can no longer be satisfied with pedagogical theories which focus exclusively on beliefs about the instructional activities of the teacher or upon the social milieu of learning.

VARIETIES OF INDEPENDENT STUDY

Since we could not find any difference in results when comparing a variety of face-to-face methods of college instruction, it makes us wonder whether any distinctive features of instruction make a difference. It is evident that independent study does seem to represent a significantly

different set of behaviors for both the student and the instructor. Indeed, the behavioral and attitudinal differences between independent study and other methods of instruction seem to be so gross as to lead to a certain expectation that distinctive consequences must result if independent study is contrasted with face-to-face methods of instruction.

The instructorphiles might expect that the instructor would produce better results when present than when absent from the teaching-learning situation. Those concerned with the "mature personality," and who see its potential realization in the college student, might believe that independent study would produce testable content-learning results superior to those produced by face-to-face instructional procedures.

There are proponents of both of these positions among college faculties. Unfortunately, these appear to be nothing but "straw man" positions, for the data support neither. With monotonous regularity, we recite the research results below. Face-to-face with their instructors, or independent of them, college students can pass their course examinations with equal facility and level of performance.

We have examined 25 studies which have presented data for comparisons of supervised independent study with all types of traditional face-to-face methods of instruction.[13] Of 81 independent comparisons presented in these studies, 40 (49.4%) favor supervised independent study and 41 (50.6%) favor face-to-face methods of instruction. The distribution of standardized differences is presented in FIGURE 4. (See page 39.) The average difference is 0.0 and the standard deviation of the differences is 1.4, indicatin gno difference in the mean examination scores of students subjected to these different methods of instruction.

The same result is found in comparisons of supervised independent study with specific face-to-face methods of instruction. For example, we located 14 studies which presented data for comparisons of supervised independent study with the lecture method. Of 50 independent comparisons made of these methods in the 14 studies, 24 (48.0%) favored supervised independent study, and 26 (52.0%) favored the lecture method. The distribution of standardized differences is given in FIGURE 5. (See page 40.) The average difference is 0.04 and the standard deviation of differences is 1.61. This mean difference is not significantly different from a "true" mean difference of zero ($t=0.17$, $P>.50$).

Again, three studies were examined which presented comparisons of supervised independent study with the discussion method. Of 7 inde-

[13] The comparisons analyzed are *not* the same ones previously analyzed examining face-to-face methods of instruction.

pendent comparisons presented in these studies, 5 (71.4%) favored supervised independent study and 2 (28.6%) favored the discussion method. The distribution of standardized differences for this comparison is given in FIGURE 6. (See page 41.) The average difference is 0.84 and the standard deviation of differences is 1.33. This average is not significantly different from a "true" mean difference of zero ($t=1.54$, $.20>P>.10$).

Finally, nine studies were examined which presented comparisons of supervised independent study with combinations of the lecture and discussion methods. Of 23 independent comparisons of these methods, 11 (47.8%) favored supervised independent study and 12 (52.2%) favored the combination methods. The distribution of differences is presented in FIGURE 7. (See page 42.) The average difference is 0.19 and the standard deviation of differences is 1.04. Here again, this average difference is not significantly different from a "true" mean difference of zero ($t=0.72$, $.50>P>.40$).

Although we again avoid presenting a discussion of comparisons of supervised independent study with the specific types of combinations of the lecture and discussion method, the data for these comparisons are presented in APPENDIX B. The same result seems to hold for each of these specific types of comparisons.

In sum, we have examined a number of comparisons of supervised independent study with face-to-face methods of teaching. The only conclusion which we can reach from these data is that neither of these methods is measurably to be preferred when evaluated by final examination performances.

THE FINAL BLOW

We still have one more chance to make a case for differences among college teaching methods. Perhaps unsupervised independent study *is* really distinctive, and its results will measurably differ from the results of other instructional procedures. The answer is "No."

Turning to the experimental comparisons, we have examined 6 studies which present comparisons of unsupervised independent study with face-to-face methods of instruction. Of 25 independent comparisons of these methods, 15 (60.0%) favor unsupervised independent study and 10 comparisons (40.0%) favor face-to-face methods of instruction. The average difference between these methods is —0.18 and the standard deviation of differences is 1.53. This average is not significantly different from a "true" mean difference of zero ($t=0.58$, $P>.50$). The distribution of standardized differences is given in FIGURE 8. (See page 43.)

We shall not belabor the point further by presenting the results of comparisons of the two types of unsupervised independent study with specific methods of face-to-face instruction. The data for these comparisons are presented in APPENDIX B. The interested reader will note that in one case the average difference is significantly different from zero. This is the comparison of self-study with the discussion method. Although there are only two independent comparisons of these methods, it should be noted that in both cases the difference favors self-study.

The last group of comparisons which we have examined is that which includes comparisons of supervised independent study with unsupervised independent study. The data for these comparisons come from 2 studies. Of 12 independent comparisons reported in these studies, 3 (25.0%) favor unsupervised independent study and 9 (75.0%) favor supervised independent study. The average difference is not significantly different from zero however, ($X=0.24$, $SD=0.79$, $t=1.02$, $.50>P>.40$). The distribution of standardized differences is presented in FIGURE 9. (See page 44.)

Here again, we shall not discuss comparisons of the two types of unsupervised independent study ("self-study," "autonomous small groups") with supervised independent study, as reported in APPENDIX B. The data for these comparisons are consistent with the findings we have reported throughout this chapter.

In a Word—Nothing

In the foregoing paragraphs we have reported the results of a reanalysis of the data from 91 comparative studies of college teaching technologies conducted between 1924 and 1965. These data demonstrate clearly and unequivocally that there is no measurable difference among truly distinctive methods of college instruction when evaluated by student performance on final examinations.

(FIGURES 1 *through* 9 *which are mentioned in this chapter appear consecutively on pages* 37-44.)

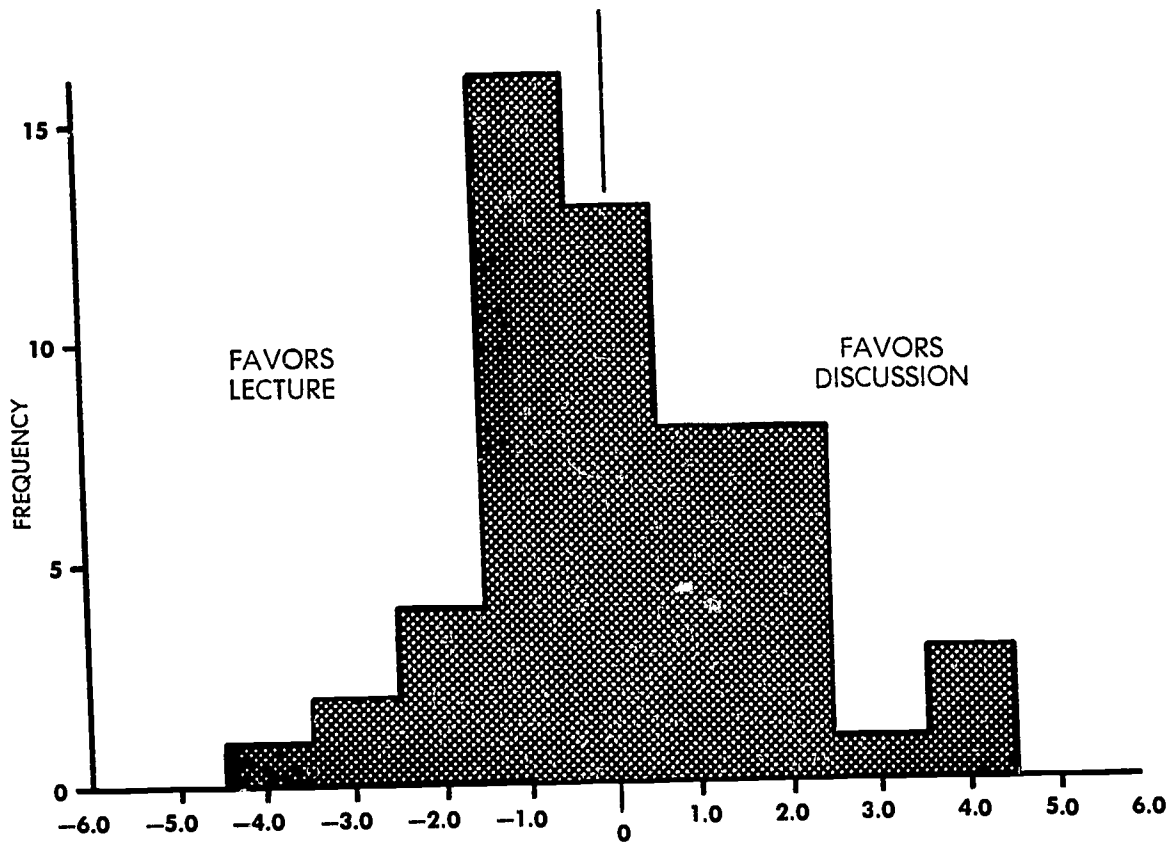
FIGURE 1

There is No Difference Between Lecture and Discussion in Face-to-Face Teaching

SIGNED DIFFERENCES*

THE DIFFERENCE	N	%
Favors Lecture	45	51.1
Favors Discussion	43	48.9
Shows No Difference	0	0.0
TOTAL	88	100.0

Studies:** 001, 004, 006, 008, 010, 011, 012, 014, 017, 020, 021, 022, 024, 025, 027, 028, 029, 030, 031, 032, 033, 034, 035, 036, 037, 039, 040, 041, 043, 044, 045, 046, 065, 086, 093, 132.



STANDARDIZED DIFFERENCES

N = 56
MEAN = 0.09
SD = 1.70
t = 0.38 P > .50

Studies:** 001, 004, 008, 011, 017, 020, 021, 025, 027, 028, 030, 032, 033, 034, 036, 037, 039, 040, 043, 044, 045, 046, 086.

* A grand total of 201 independent and non-independent comparisons were found for which 54.7% favor lecture, 44.8% favor discussion and 0.5% show no difference.
** The full citations as to source are identified by study number in the bibliography.

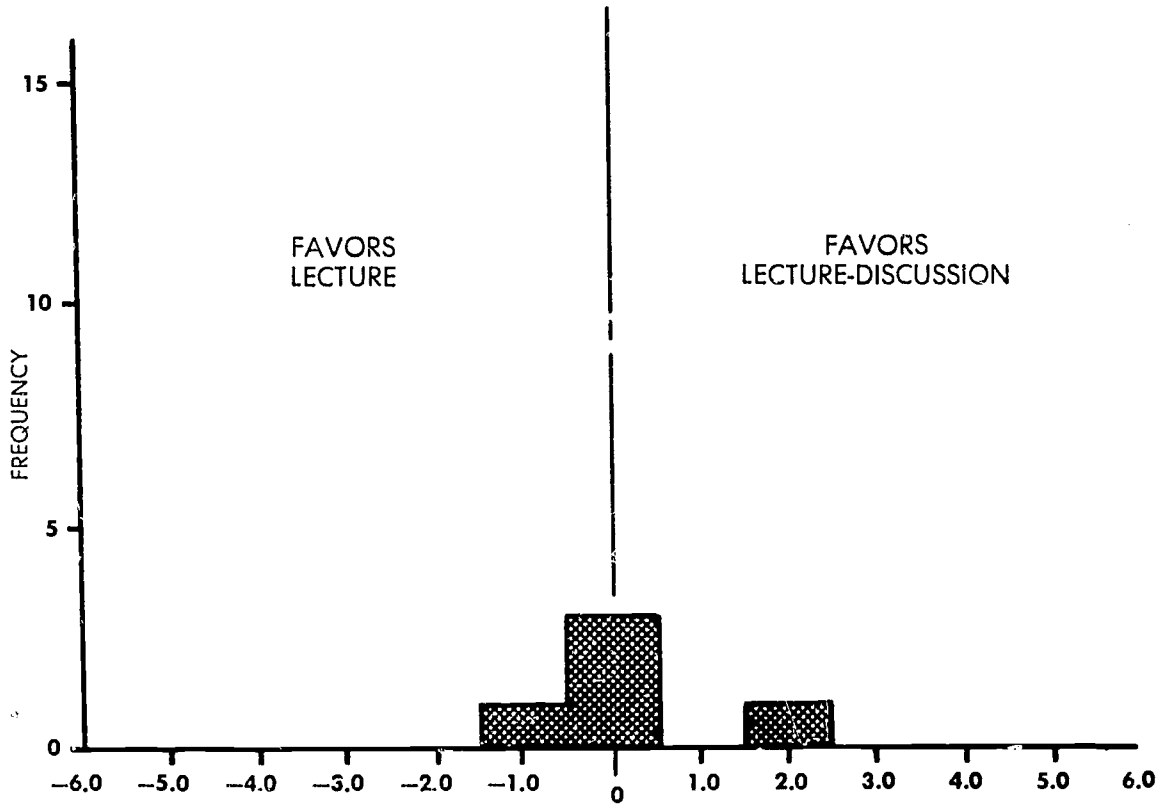
FIGURE 2

Lecturing or Lecturing and Discussing Make No Difference in Face-to-Face Teaching

SIGNED DIFFERENCES*

THE DIFFERENCE	N	%
Favors Lecture	3	37.5
Favors Lecture-Discussion	4	50.0
Shows No Difference	1	12.5
TOTAL	8	100.0

Studies: 002, 003, 018, 019, 048, 062, 087.



STANDARDIZED DIFFERENCES

N = 5
MEAN = 0.19
SD = 0.84
t = 0.46 P > .50

Studies: 002, 003, 018, 048, 062.

* A grand total of 59 independent and non-independent comparisons were found for which 45.8% favor lecture, 52.5% favor lecture-discussion and 1.7% show no difference.

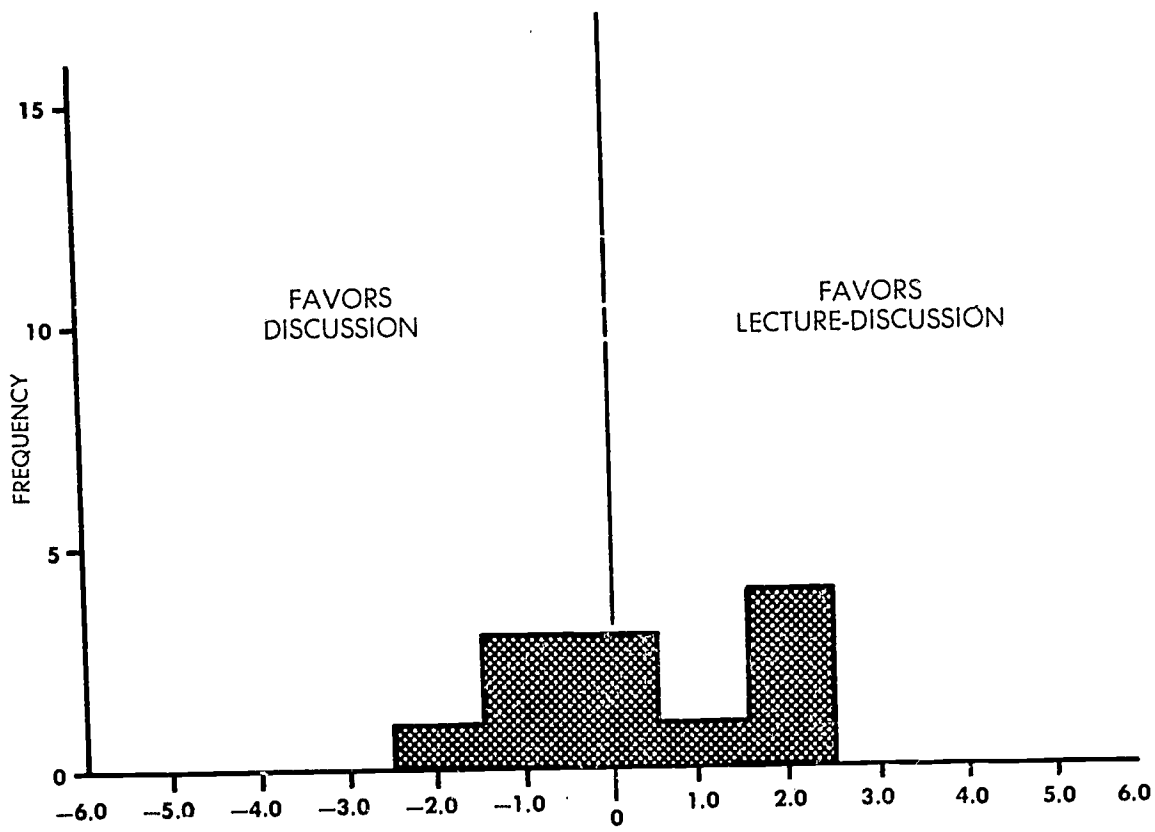
FIGURE 3

Discussion Makes No Difference Compared to Lecture and Discussion in Face-to-Face Teaching

SIGNED DIFFERENCES*

THE DIFFERENCE	N	%
Favors Discussion	5	41.7
Favors Lecture-Discussion	5	41.7
Shows No Difference	2	16.6
TOTAL	12	100.0

Studies: 009, 026, 055.



STANDARDIZED DIFFERENCES

N = 12
MEAN = 0.20
SD = 1.53
t = 0.43 P > .50

Studies: 009, 026, 055.

* A grand total of 16 independent and non-independent comparisons were found for which 50.0% favor discussion, 31.2% favor lecture-discussion and 18.8% show no difference.

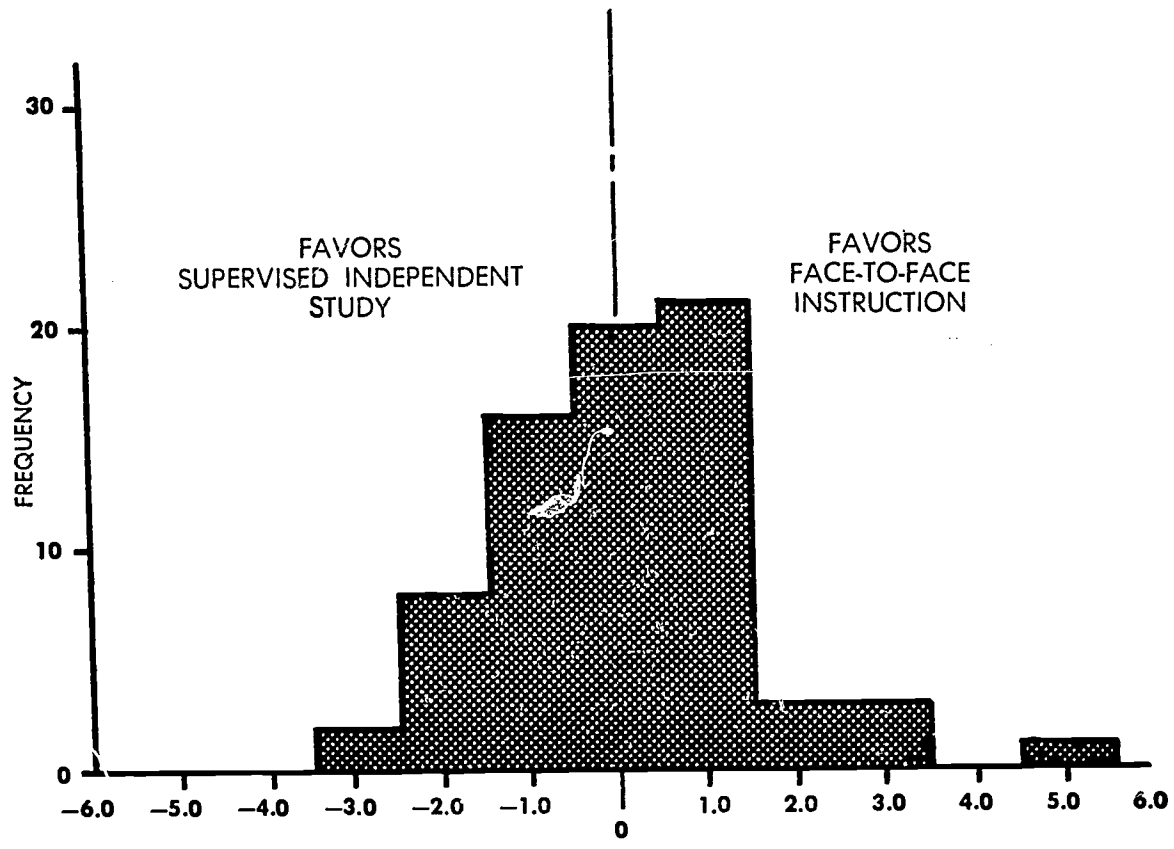
FIGURE 4

Supervised Independent Study and Face-to-Face Instruction Produce the Same Results

SIGNED DIFFERENCES*

THE DIFFERENCE	N	%
Favors Supervised Independent Study	40	49.4
Favors Face-to-Face Instruction	41	50.6
Shows No Difference	0	0.0
TOTAL	81	100.0

Studies: 049, 050, 052, 053, 055, 058, 059, 060, 061, 066, 068, 069, 071, 073, 075, 076, 077, 078, 079, 080, 082, 083, 084, 086, 088.



STANDARDIZED DIFFERENCES

N = 74
MEAN = 0.0
SD = 1.47
t = 0.01 P > .50

Studies: 050, 052, 053, 055, 058, 059, 060, 061, 066, 068, 069, 071, 073, 075, 076, 077, 078, 079, 080, 084, 086, 088.

* A grand total of 116 independent and non-independent comparisons were found for which 50.0% favor supervised independent study and 50.0% favor face-to-face instruction.

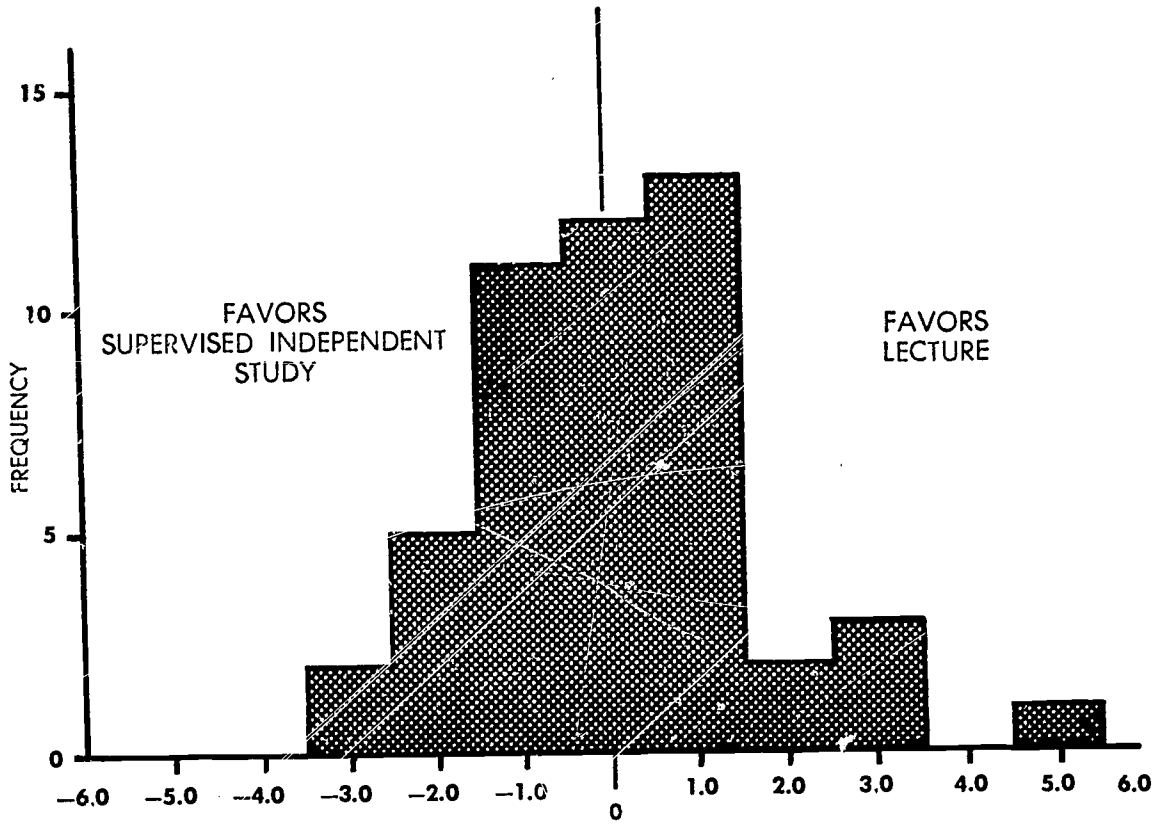
FIGURE 5

Supervised Independent Study and Lecturing Are Not Different in Examination Results

SIGNED DIFFERENCES*

THE DIFFERENCE	N	%
Favors Supervised Independent Study	24	48.0
Favors Lecture	26	52.0
Shows No Difference	0	0.0
TOTAL	50	100.0

Studies: 049, 050, 052, 058, 059, 060, 068, 071, 073, 075, 077, 079, 086, 088.



STANDARDIZED DIFFERENCES

N = 49
MEAN = 0.04
SD = 1.61
t = 0.17 P > .50

Studies: 050, 052, 058, 059, 060, 068, 071, 073, 075, 077, 079, 086, 088.

* A grand total of 72 independent and non-independent comparisons were found for which 47.2% favor supervised independent study and 52.8% favor lecture.

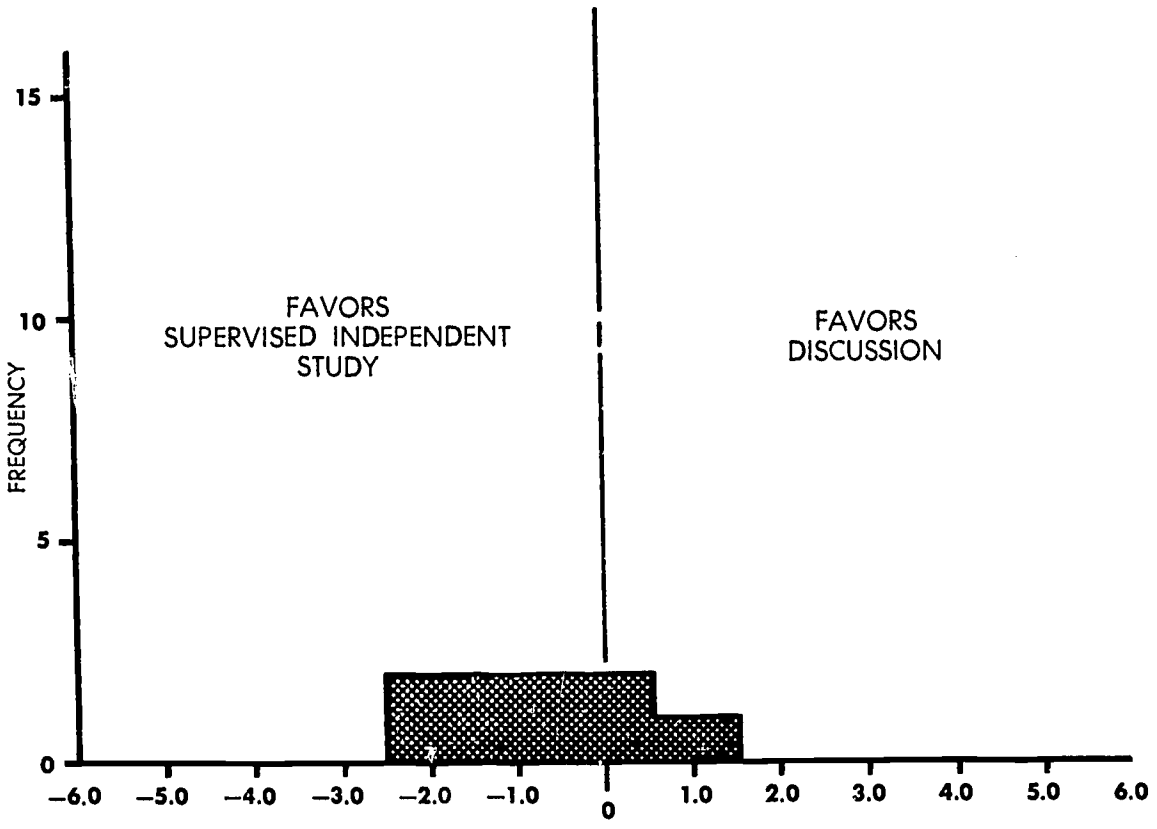
FIGURE 6

Supervised Independent Study and Discussion Method Produce No Differences

SIGNED DIFFERENCES*

THE DIFFERENCE	N	%
Favors Supervised Independent Study	5	71.4
Favors Discussion	2	28.6
Shows No Difference	0	0.0
TOTAL	7	100.0

Studies: 053, 055, 076.



STANDARDIZED DIFFERENCES

N = 7
MEAN = −0.84
SD = 1.33
t = 1.54 .20 > P > .10

Studies: 053, 055, 076.

* A grand total of 9 independent and non-independent comparisons were found for which 77.8% favor supervised independent study and 22.2% favor discussion.

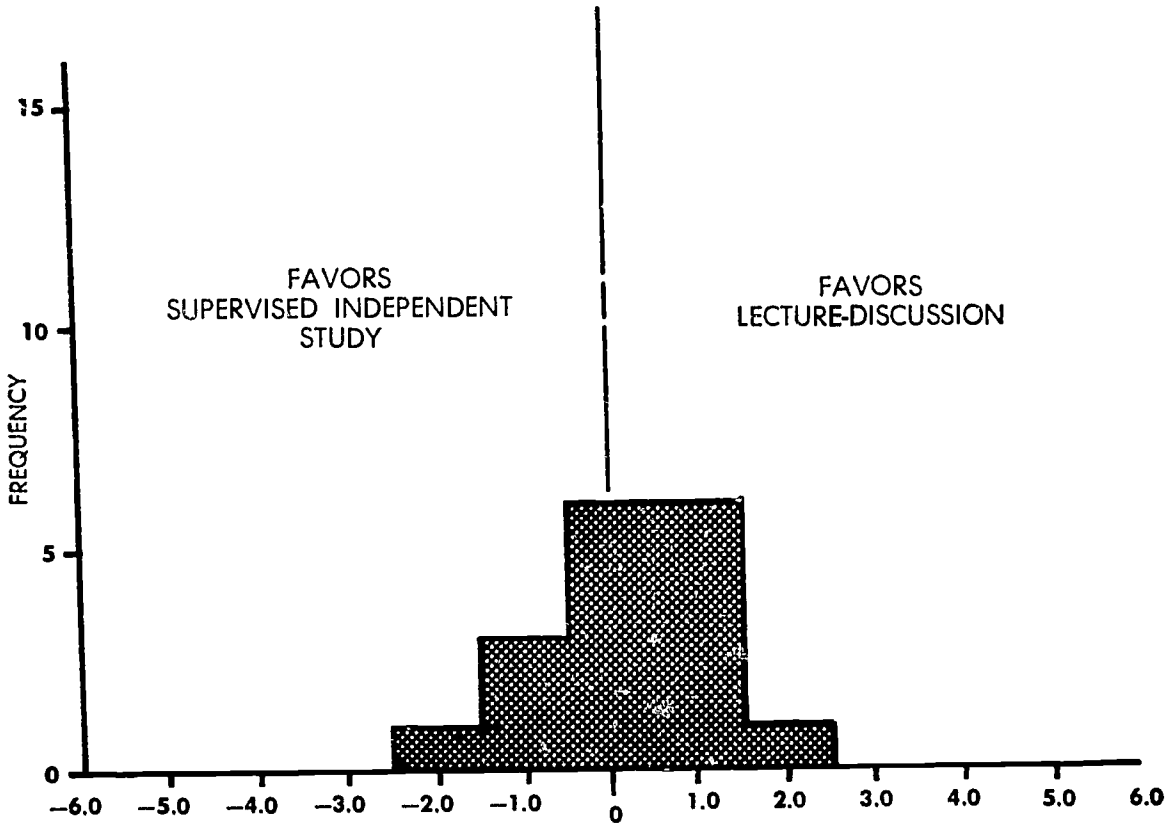
FIGURE 7

Supervised Independent Study and Lecture-Discussion Instruction Produce No Differences in Examination Scores

SIGNED DIFFERENCES*

THE DIFFERENCE	N	%
Favors Supervised Independent Study	11	47.8
Favors Lecture-Discussion	12	52.2
Shows No Difference	0	0.0
TOTAL	23	100.0

Studies: 055, 061, 066, 069, 078, 080, 082, 083, 084.



N = 17
MEAN = 0.19
SD = 1.04
t = 0.72 .50 > P > .40

Studies: 055, 061, 066, 069, 078, 080, 084.

* A grand total of 34 independent and non-independent comparisons were found for which 50.0% favor supervised independent study and 50.0% favor lecture-discussion.

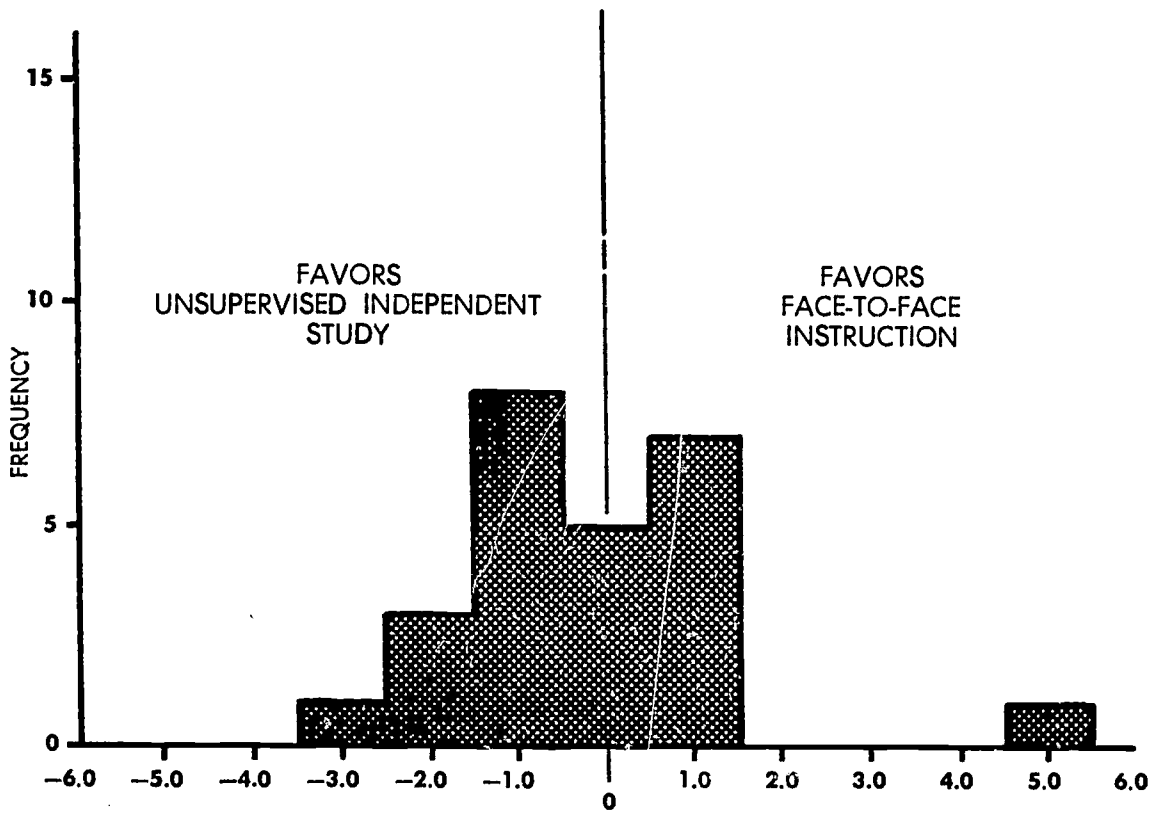
FIGURE 8

Even Unsupervised Independent Study Is No Different From Face-to-Face Instruction

SIGNED DIFFERENCES*

THE DIFFERENCE	N	%
Favors Unsupervised Independent Study	15	60.0
Favors Face-to-Face Instruction	10	40.0
Shows No Difference	0	0.0
TOTAL	25	100.0

Studies: 004, 036, 056, 059, 067, 073.



STANDARDIZED DIFFERENCES

N = 25
MEAN = −0.18
SD = 1.53
t = 0.58 P > .50

Studies: 004, 036, 056, 059, 067, 073.

* A grand total of 31 independent and non-independent comparisons were found for which 58.1% favor unsupervised independent study and 41.9% favor face-to-face instruction.

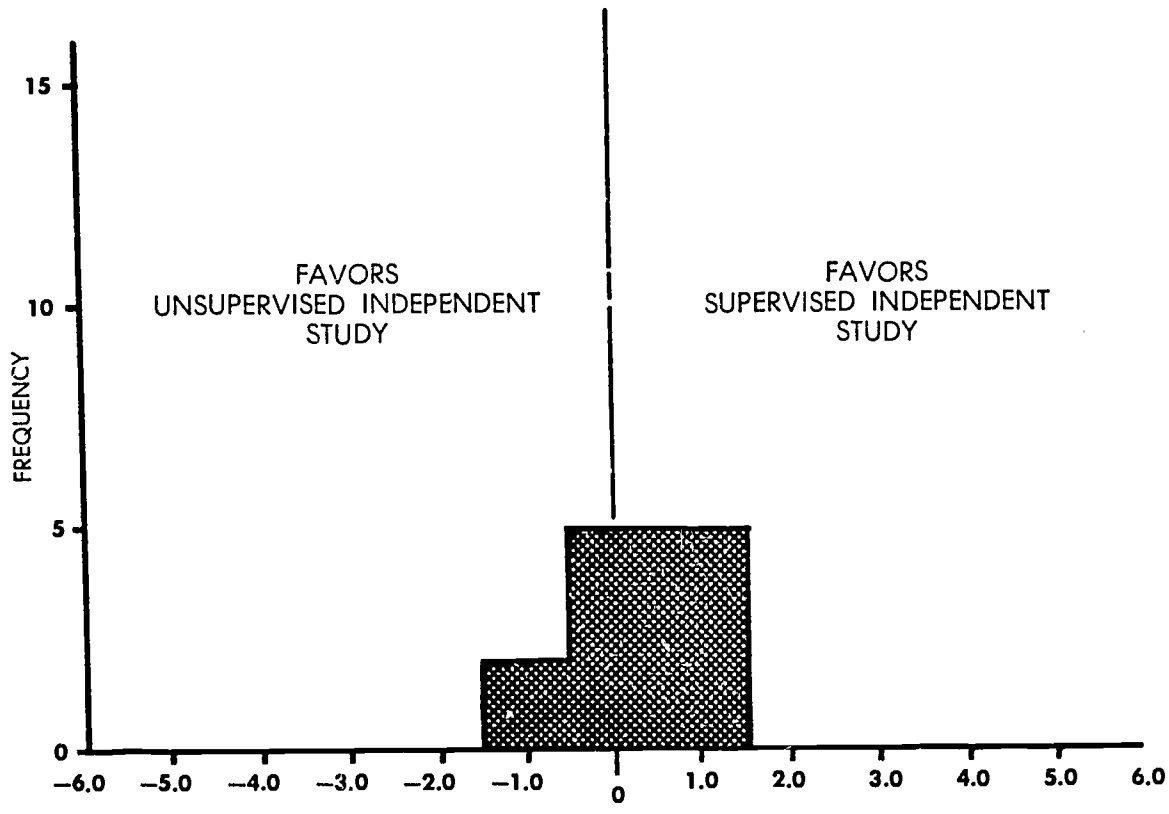
FIGURE 9

Independent Study, Supervised or Unsupervised, Produces No Differences

SIGNED DIFFERENCES*

THE DIFFERENCE	N	%
Favors Unsupervised Independent Study	3	25.0
Favors Supervised Independent Study	9	75.0
Shows No Difference	0	0.0
TOTAL	12	100.0

Studies: 004, 059.



STANDARDIZED DIFFERENCES

N = 12
MEAN = 0.24
SD = 0.79
t = 1.02 .50 > P > .40

Studies: 004, 059.

* A grand total of 12 independent and non-independent comparisons were found for which 25.0% favor unsupervised independent study and 75.0% favor supervised independent study.

CHAPTER
4

The Past Is Not Prologue

It will add nothing to knowledge to continue to do in the future what researchers have done in the past in studying comparative college teaching methods. We are confident that to follow the example of past researchers will only duplicate their results, and nothing more.

In this monograph we have reported the results of a careful and systematic reanalysis of the data of almost 100 comparative studies of different college teaching methods. We have found no shred of evidence to indicate any basis for preferring one teaching method over another as measured by the performance of students on course examinations.

New Directions

Research on comparative college teaching methods has to move in new directions and the new directions in which research must move are revealed in our findings.

The evidence is crystal clear that there must be something in common among all college teaching methods. Past research has taken as its starting point a belief or faith that a preferred method of college instruction is somehow or other superior to alternatives. We think that the new faith which generates the future research must be grounded in the empirical

conclusion that *there is something in common,* or shared, among all college teaching methods.

COMMONALITIES AMONG TEACHING METHODS

Future research on comparative teaching methods must focus on the question: "What is there that is the same about any two different teaching methods?"

It seems very clear that while the technologies differ—a lecture without discussion is technically a very distinctive teaching situation from independent study, for example—these obvious technical differences do not make any difference in student performance. What is visible as a difference in technology is not a significant difference for the teaching-learning situation.

The first task, then, is to find out what is common to the very diverse teaching methods employed in colleges. Once we have answered that question we may then be in a position to analyze the consequences of differences among teaching methods.

At this point it seems obvious that an alternative question could be posed: "Have the wrong outcomes been measured, therefore making the research really irrelevant?" This is a beguiling alternative and worthy of attention.

The contemporary world is described as undergoing a *knowledge explosion.* We usually mean by knowledge a body of fact, and the models or ideas that man employs to make sense out of these facts. With the contemporary acceleration in the growth of knowledge, every citizen faces the continuing problem of remaining current in his knowledge. The colleges and universities of the country are central institutions in developing new knowledge through research, and transmitting the accumulated knowledge to the next generation through teaching.

Given, then, the knowledge explosion, it is difficult to make sense out of downgrading the transmission of knowledge as a primary function of college and university education. It is gratuitous to sneer at content learning as the preoccupation primarily of the vocationally-committed students. All citizens need to have an adequate grasp about the world in which they live, and the sense that men have made out of this complexity. This is content in the best, and indeed, only sense in which the term may be applied.

Thus, to say that content learning, as measured by course examinations, is not relevant to the reasons why students are in college is simply to fly in the face of reality. Students are in institutions of higher educa-

tion to learn content. We measure this learning by examinations which are content-oriented. We seem to be measuring an outcome which is relevant to one of the major functions of an institution of higher education.

All researchers have been inclined to overlook what is, perhaps, the outstanding commonality among teaching methods compared in a given study. This is the textbook(s) utilized. As Professor Hilgard so perceptively stated:

> Most studies have relied very heavily on a common textbook in all the courses, and, in order to be "fair" most of the examination questions are based on that book. I can't help but believe that more careful exposition goes into a good textbook than a lecturer can put into a lecture. . . . Hence I believe we are often measuring what the student learned from his textbook, which makes it quite indifferent what amusing stories the lecturer told, or how skilled the student was in winning a point in a class argument. The objections to the textbook are something like the objections to the lecture, but where there is really something to be learned (like in an anatomy course) everyone would recommend a textbook. Maybe textbook's aren't so bad, after all, but in any case they may be so *powerful* as to override differences in teaching.[1]

It may very well be that the most pervasive commonality among teaching methods is the employment of and dependence on textbooks and other reading materials. Perhaps the "no difference" results of comparing teaching methods can be attributed largely to the powerful impact of textbooks which cannot be washed out by any known methods of instruction! More important, however, if textbooks are the most significant influence in student learning, then future studies should examine differences among textbooks rather than among teaching methods.

Other aspects of the consequences of college education for students have been studied by scholars like Newcomb and Sanford. Newcomb, for example, has found that the attitudinal climate of a college faculty influences the attitudes of students toward conformity with the faculty climate of attitudes: a liberal faculty will produce liberally-oriented students, in spite of their conservative upbringing. In a more classical tradition, Hutchins and others have argued that higher education should teach people to think. Although no one would contend that thinking is a low-order outcome of education, thinking needs content and the skills of thought must be applied to subject matter.

It would be unfortunate if, because subject matter-oriented examina-

[1] Personal communication from Professor Ernest R. Hilgard in response to a request to comment on a pre-publication draft of this monograph. We are indebted to Professor Hilgard for this important point and other observations useful in improving this monograph.

tions proved no differences among teaching methods, we would then conclude that subject-oriented examinations do not measure useful, important, or relevant features of the impact of higher education upon students. It works the other way around: the content retention among students is the same regardless of the methods by which they are taught. This content retention is important as a way of staying on top of the knowledge explosion.

The commonality among disparate college teaching methods lies in their equal facility for transmitting knowledge to the next generation. This is a primary function of higher education and it seems to be fulfilled in a variety of ways with equal effectiveness.

MODELS OF TEACHING-LEARNING

A second new direction for research is to raise the entire issue of how we model teaching-learning. It seems clear that our knowledge of pedagogy, although distinctively clear regarding its differential technologies, is incapable of predicting differences between distinctive technologies. Most theories of learning simply ignore the pedagogical side. We are now very much in need of models of teaching-learning. The Skinnerian model and the technology of operant conditioning move in exactly the right direction because in these is a self-conscious concern with the teaching-learning linkages.

We may then summarize the new directions for research as:

(1) To find the commonalities among distinctive technologies of college teaching; and

(2) To develop models of the teaching-learning situation.

These two goals are obviously related. Knowing what is common to all teaching methods will provide better bases for modeling the teaching half of the teaching-learning situation. We have suggested in CHAPTER 1 some of the considerations that might very well bear upon the learning half of the teaching-learning situation.

What About Here and Now

There are some obvious educational policy implications to be drawn from the results of this monograph and its three companion monographs.

Let us start with some pretty obvious assumptions:

(1) Enrollments in post high school education in the United States will continue to rise in absolute numbers. This is true (a) because of

the age distribution of the population and (b) because of the social policies which encouraged continued education beyond high school levels.

(2) The number of teachers available for post-high school instruction will not keep pace with the growing demand because: (a) The age distribution of present teachers reflects the small number of students trained for teaching during the depression and World War II; (b) There are too few students presently in training for positions as future teachers; and (c) There are alternative career opportunities for those now in training to pursue research or administration rather than teaching within institutions of higher education, or to follow careers in government and business.

(3) There will be a continuing and accelerated crisis in the financing of higher education especially since the proportion of all students in post high school education is increasing far more rapidly in publicly supported institutions than in private ones.

The policy implications of these assumptions and the facts set forth in this monograph are very clear.

COST-BENEFIT ANALYSIS

Increasing attention will be demanded of college and university administrators to the cost-benefit analysis of various teaching methods. Up to this point, the "benefit" portion of cost-benefit analysis has largely depended upon private opinion and prejudice. We think that we have demonstrated in this monograph that the usual prejudices regarding preferred college teaching methods are no longer acceptable as bases for alleging the benefits of particular teaching technologies.

Indeed, since there are no differences among a wide range of teaching technologies we may assume that their respective benefits are equal. This, then, turns the attention in cost-benefit analysis to the cost side of the issue.

In making the costing decisions the obvious strategy would seem to be to pay out as little as possible for instructional costs. As we have pointed out in a companion monograph, the subrosa means employed up to now to minimize instructional costs has been to use low cost graduate teaching assistants in undergraduate instruction, and especially in lower division instruction.[2] The more visible means for powering per student instructional costs has been to increase the size of individual classes to markedly raise the student-teacher ratio. Large lecture classes

[2] See Robert Dubin, and Fredric Beisse, *The Academic Underworld: Graduate Assistants* (Eugene, Oregon: CASEA, 1969).

have become a common feature of undergraduate instruction and represent a very rational response to improving the cost-benefit ratios of instruction.

A distinction needs to be drawn between short term cost savings and long term cost savings in college instruction. The contemporary cost saving practices of employing large classes and using cheap graduate assistant instructors are short term cost solutions. There are available teaching technologies which have long term cost saving features that may be more beneficial in the long run but which require high level initial investments. Thus, computer-assisted instruction requires high initial investment costs in computer consoles and software, such as teaching programs. Over the long run these costs may spread very rapidly over the student population and result ultimately in a much lower per student cost of instruction. At a somewhat lower initial investment educational television may be installed and with the use of video-tapes could result in long term per student cost reductions.

RADICAL INNOVATIONS

The instructional method alternatives so far considered in terms of cost benefit analysis all assume a campus installation with either a resident and/or a commuter student body. More radical innovations might involve dispensing with the campus entirely or modifying it in major dimensions. If self-study and educational TV are as good as face-to-face contact in promoting content learning, then the need for students to confront their instructors physically is materially reduced. Given the modern technologies of long distance and instantaneous transmission of audio and visual signals, it is entirely within the realm of possibility that the traditional school and campus will no longer be the locale of instruction. Doctors and dentists now receive updating of their knowledge in their home locality by transmission of medical information in television and other kinds of distance transmission methods. The University of the Air has already demonstrated how school teachers as well as college students can gain course credits without leaving their homes.

Whether the academic administrator likes it or not, and since he is usually an ex-college professor he will not like it, these academic decision-makers are going to be faced with the need for applying cost-benefit analysis to instructional costs. It is reasonable to expect that if there are no measurable differences in content learning among various teaching methods that the least costly will be adopted.

There are obvious secondary costs involved in present operations in

colleges and universities which are related to the modes of instruction employed. With radical changes in modes of instruction that are no longer either time-bound or place-bound it could very well be that the secondary costs could be even more drastically reduced than the primary costs of instruction. Thus, for example, the 12, 18, or even 24 hour-a-day period of instruction could well replace the present limited period for scheduled classes.

OTHER EDUCATIONAL GOALS

The policy implications relating to other goals of education, distinct from content learning, may also require re-examination. The usual justification for public education is that, first, it prepares the citizen for useful contribution to the society, usually in a vocational sense. Its second justification lies in the belief that the educated citizen is more responsible and constructive in his citizenship than the uneducated citizen. Content learning in subject matter courses clearly contributes to the first goal. If the second justification were to achieve a greater emphasis than it does at present, a whole new set of activities may be brought into the college environment. These activities may have to do with the modification or development of attitudes; the growth and development of emotional life and expressive behavior; or even "mind-expanding" beyond the present limits of conscious, and self-conscious behavior and mental life. Except for very limited knowledge of "brain-washing" and the reports of travellers on psychedelic trips, few administrators and members of present college faculties, except psychiatrists, are trained or knowledgeable about providing these kinds of educational experiences.

The Launching Pad

We set out in this monograph to provide a launching pad for new directions in research on comparative college teaching methods as well as to make clear the grounds for educational policy decisions regarding college teaching methods.

The data analyzed in this study make clear that nothing new will be discovered about college teaching methods until we ask new questions and seek their answers in research which departs significantly from that pursued in the past.

The major academic policy consequence of this study is to change the stance of academic decision-makers from one of apologist, to one of confident realist in evaluating the cost benefits of various college teaching methods. In the past, college administrators have apologized as they

made their painful decisions favoring mass methods of instruction, sometimes even employing low quality instructors. The burden of the evidence suggests that at least the methods of instruction are not germane to such decisions, although we hope that the quality issue will remain paramount for the policy decision-makers.

The technical appendix which follows will be important to those who want to know how exactly we arrived at our conclusions. APPENDIX B contains supplementary data upon which these conclusions rest.

APPENDIX
A

Methods of Analysis

In libraries, research centers and personal files across the country, great amounts of data are available today concerning everything from voting behavior to migration of college students. Most of these data, however, were gathered in independent research projects conducted in different years, at different places, and with different methods. This suggests two immediate questions for secondary analysis: (1) Can we in some way combine these data? (2) Given that techniques can be devised for combining various segments of the data, what limitations are placed on conclusions drawn from the combined data?

In responding to these questions, first it must be recognized that there is no one method applicable to any and all problems of secondary analysis. Each research problem is quite unique. The methods of analysis to be employed and the types of conclusions reached must necessarily be geared to the specific research problem. Second, it should be noted that the questions raised above are not distinct. On the contrary, conclusions and inferences based upon secondary analysis will be influenced, to a great extent, by the methods and techniques used to analyze these data. Finally, the problem of re-using data to test ideas is a relatively new one to the social sciences. Although the practice of adding up data on a subject to test theories, or build new theories, is a time-honored procedure in the natural sciences, only in the last few decades have large amounts of data begun to accumulate in the social sciences. As a result, the user of

other people's data in the social sciences will find little in the literature or in the past experiences of his peers to help him solve his specific research problems. Furthermore, the data accumulation process is rapidly expanding and the problems faced by users of secondary data are likely to become ever more pressing.

In the following pages we would like to contribute what we might to the solution of the problems briefly pointed out above by discussing the methods of analysis which we have used in reanalyzing a large amount of data on methods of instruction in colleges and universities. The question which we have attempted to answer in reanalyzing these data is: Given a population of seven million adults or near-adults attending two- and four-year colleges and universities as students, what can we say about the relation among various methods for instructing them and the outcomes produced when measured on final examinations in their courses? Although the methods of analysis that we have decided upon are geared to this specific question, perhaps the reader can glean something from the problems which we have faced and the assumptions which we have had to make.

Source and Nature of Data

After having decided upon our particular research question, the first problem encountered involved deciding what resources we were going to use in analyzing the relative worth of different methods of instruction. We were aware that there existed a large number of empirical studies of different methods of teaching and that previous summaries of this research literature had utilized authors' conclusions and had not, to our knowledge rigorously examined the data upon which these conclusions were based. (See *Bibliography*) Also, recognizing that researchers sometimes mistakenly interpret their own data and reach faulty conclusions, we felt that if we were to make a unique contribution in this area, we must attempt to add up the *data* of the field, rather than the conclusions of the researchers who had marshaled the data.

Thus, we collected 91 studies reporting one or more experimental comparisons of different teaching technologies. (See *Bibliography*) The studies that we examined appeared in the literature between 1924 and 1965 as journal articles or doctoral dissertations. For the reader's convenience, we have listed in TABLE 1 the sources of the studies we utilized.

Before deciding upon the specific methods of analysis to be employed and the statistical techniques we would use, we had to examine carefully the material we had collected. Our analysis, of course, would be influenced

Table 1. Sources Utilized: Experimental Studies of Teaching Methods.*

Journal	Issue
Dissertation Abstracts	1-27
American Journal of Physics	20
American Psychologist	4, 8
British Journal of Educational Psychology	35
California Journal of Educational Research	13
College and University	26
Education	46
Educational Administration and Supervision	14, 24, 39
Genetic Psychological Monographs	4
Improving College and University Teaching	10-13
Journal of Applied Psychology	13, 16, 36, 38
Journal of Chemical Education	9
Journal of Colorado-Wyoming Academy of Science	4
Journal of Education	27
Journal of Educational Psychology	1-56
Journal of Educational Research	9, 22, 29-43, 45-49, 55, 57-59
Journal of Engineering Education	36
Journal of Experimental Education	12-13, 20-27
Journal of Higher Education	2, 10-12, 24, 26, 35
Journal of Medical Education	31
Journal of Psychology	31
Journal of Social Psychology	43
Phi Delta Kappan	12-14, 46
Psychological Bulletin	52
Psychological Monographs	45, 65-68, 71-72
School and Society	21, 25-26, 80
School Review	34
Science Education	11-21, 26-28, 41-43
Social Education	20
University of Oregon Publications (Educational Series I)	1

*This list does not include institutional research reports received by our office.

by whether researchers had used comparable methods in doing their research and would depend upon whether they presented comparable data. We found that 74 of the 91 studies we had collected were comparable with respect to research design and the data presented on student examination performances.

RESEARCH DESIGN

The most common research design encountered involved one or more experimental groups taught by one method of instruction and one or more control groups taught by another method of instruction. Both groups were made up of students who had enrolled in the same course and who had been arbitrarily assigned either to the experimental or control section. In many studies, these groups were tested prior to the outset of the experiment in order to insure their comparability on such attributes as age, sex, intelligence, aptitude, etc. In a few studies, the

groups were given pre-tests on the course content in order to enable analysis of gains in knowledge at the end of the experimental period. Usually at the end of the term or semester in which the experiment took place, both groups were administered final examinations covering course content. The average test scores (or average gain scores, if available) were then compared and the results of the comparison were given as evidence of the relative effectiveness of the methods of instruction being compared. If the score of a lecture group, for example, was higher than that of a discussion group, then, given the conditions of the particular study, the lecture method was judged to be superior to the discussion method.

Obviously, this type of experiment only approximates the ideal experimental situation. However, as Campbell and Stanley have noted, this type of design, ". . . should be recognized as well worth using in many instances in which . . . [true experimental designs] . . . are impossible."[1]

DATA PRESENTED

As a minimum, the 74 studies that we found to be comparable in general procedure presented descriptions of the teaching methods compared and examination scores for the experimental and control groups on course content examinations. In many studies, the data used to equate the groups were presented, and in a few studies reliability data were given for the course content tests administered. Also, almost every study presented data on one or more variables used in addition to student performance for measuring outcomes of the teaching-learning black box; for example, student attitudes towards the course or instructor, or social adjustment.[2] We were primarily interested, however, in the descriptions of the methods of teaching being compared and in the data presented on the criterion variable—student performance on course content examinations.

In CHAPTER 3, we have briefly described the general methods of teaching that were compared and the central ideas and assumptions concerning the teaching-learning linkage upon which these methods were based. We shall not repeat that discussion here. Suffice it to say that there was little systematic attempt among the studies utilized to measure the actual

[1] D. T. Campbell, and J. C. Stanley, "Experimental and quasi-experimental designs for research on teaching," in N. L. Gage (Ed.), *Handbook of Research on Teaching*, (Chicago: Rand McNally, 1963), p. 217.

[2] It should be noted, however, that there was little attempt to replicate the results on these variables across different studies. As a result, it would be impossible to summarize the data for these other outcomes in any systematic manner.

differences between teaching methods compared. For the most part, these differences were described by word labels employed by the researchers. In several instances we classified teaching methods on the basis of descriptions contained in the research report.

The significant point, however, is that although there apparently was a great deal of individual variation within particular studies and between studies in terms of the specific methods of teaching employed, and although the labels attached to these specific methods varied tremendously between studies, on the basis of the descriptions presented we were able to identify and define a small number of general methods of teaching. These are the general methods of teaching that are described in CHAPTER 3 and that are the basis for combining the results from different studies.

The data on performance were somewhat more precise. The data fell into two categories. In 18 of the 74 studies that we found to be comparable, the author only reported the mean performance scores or mean gain scores of groups taught by different methods. In the remaining 56 studies, however, the author not only reported the average performance scores, but also the sample sizes and standard deviations of the respective groups. These data, to a large degree, dictated the statistical procedures used to analyze the data.

COMBINING RESULTS OF DIFFERENT STUDIES

At this point, the reader is perhaps worried by the variability which we have alluded to in the research studies used. The obvious question is: Doesn't such variability preclude cumulating the data from these studies?

With respect to the variability observed in experimental procedures, we agree wholeheartedly with Campbell and Stanley who have noted that,

> . . . the more numerous and independent the ways in which the experimental effect is demonstrated, the less numerous and less plausible any singular rival invalidating hypothesis becomes. The appeal is to parsimony. The "validity" of the experiment becomes one of the relative credibility of rival theories: the theory that X had an effect versus the theories of causation involving the uncontrolled factors. If several sets of differences can all be explained by the single hypothesis that X has an effect while several separate uncontrolled-variable effects must be hypothesized, a different one for each observed difference, then the effect of X becomes the most tenable. This mode of inference is frequently appealed to when scientists summarize literature lacking in perfectly controlled experiments. Thus Watson (1959, p. 296) found the evidence for the deleterious effects to maternal deprivation confirmatory because it is supported by a wide variety of evidence-types, the specific inadequacies of which vary from study to study. Thus, Glickman (1961), in spite of the presence

of plausible rival hypotheses in each available study, found the evidence for a consolidation process impressive just because the plausible rival hypothesis is different from study to study.[3]

Similarly, with respect to the measurement and description of the teaching methods compared and multiple course content examinations, Webb and others have noted that, "... When multiple operations provide consistent results, the possibility of slippage between conceptual definition and operational specification is diminished greatly."[4]

Thus, the problem of whether or not we can combine the data from numerous and independent studies is really not in question. The conclusions which we reached are, however. If our results are inconsistent, then we shall have to treat systematically each of the plausible rival hypotheses which might account for this inconsistency. However, to the extent that our results are consistent, we can have confidence that the observed differences, or lack thereof, are tenable.

Statistical Procedures

In choosing the statistical procedures which were used to analyze the data which had been gathered, we were not only guided by the limitations of this data, but also by the following considerations:

(1) We wanted to make full utilization of the examined data; and

(2) We wanted to use statistical techniques that would give us the greatest precision possible.

As a result, we decided to use two different sets of statistical procedures: an analysis of signed differences in mean group performance; and, an analysis of standardized differences in mean group performance.

USING ALL THE DATA

Earlier we noted that in 18 of the 74 studies found to be comparable, only mean scores or mean-gain scores were presented for experimental and control groups on a course content examination. However, the accompanying standard deviations and sample sizes were not presented. In order to make use of the comparative information contained in these studies, we therefore had to utilize a statistical procedure which was based on the average scores only. The only procedure which we felt met this restriction was a sign test of the difference in mean achievement. The procedure involved computing a difference in mean achievement for

[3] Campbell and Stanley, *op. cit.*, p. 206.
[4] E. J. Webb, et al., *Unobtrusive Measures* (Chicago: Rand McNally, 1966), p.5.

each comparison, and to record whether this difference, however small, favored method 1 or method 2.

This signed-difference was computed for every comparison in the 74 studies presenting comparable data and the results were combined for each type of comparison. For example, in comparisons of the lecture and discussion methods of teaching, it was found that 43 of the reported differences in average group performance favored the discussion method, while 45 of the differences favored the lecture method. The combined results for each comparison of teaching methods, independent comparisons only, are presented in the upper boxed section of each of the figures in the text. The same sign test for *all* comparisons between each pair of teaching methods contrasted is given in the footnote to each figure.

The reader should note that although these procedures enable us to make use of the data from the greatest number of experimental comparisons examined, they are relatively imprecise and, in and of themselves, might support unwarranted conclusions. For example, although approximately the same number of reported comparisons favored the lecture and discussion methods, it is possible that the comparisons favoring the discussion method are highly significant whereas the comparisons favoring the lecture method are relatively insignificant. The conclusion we would want to reach if this were, in fact, true would be quite different from the conclusion based solely on the sign test. In the former case, we would want to conclude that the discussion method was superior, whereas in the latter case we would emphasize that the two methods were equally effective.

The point is, the sign test does not give us any indication of the magnitude or distribution of differences between any two methods of teaching. Thus, although we had to sacrifice several comparisons, we felt it not only desirable, but also necessary to have a more precise means of measuring the differences in average group performances.

INCREASING PRECISION

In 56 of the 74 studies the author not only presented average performance scores, but also the accompanying sample sizes and standard deviations of each group compared. With these data in hand, we were in a position to compute standardized scores to represent the difference in group performance for each comparison. That is, we would represent the difference reported in each comparison in standard units which could then be easily compared.

The statistical method of standardizing each difference involved divid-

ing the observed difference by an estimate of the standard error of the distribution of that difference, symbolically:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}}$$

where, $$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{N_1 - 1} + \frac{s_2^2}{N_2 - 1}}$$

and $$s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{N - 1}$$

This statistic is the appropriate test of the significance of difference between means when population variances are not known but can be assumed to be equal.[5]

We utilized the value of t in combining the results of comparisons across different studies. We also used this statistic to measure the significance of the difference in group performance for each individual comparison.[6]

The most important function that this statistic served was in giving us a picture of the combined results for each type of comparison. After the t-ratios had been computed separately for each comparison, a histogram of t-values was created. This histogram is presented for each type of teaching method comparison in the figures of the text for all independent comparisons.

The fact that we are distributing t-values should in no way influence the shape of the resulting distributions. Thus, there is no reason to expect, *a priori*, that the resulting distributions will be unimodal, bimodal, etc. However, the way in which the observed differences (as measured by t-ratios) do distribute themselves will greatly influence the types of conclusions we might reach on the basis of each distribution. For example, if a distribution turns out to be bimodal, we will conclude that there is some factor or factors operating systematically to produce this result, and will proceed with some sort of control analysis to discover what this factor is.

[5] This ratio will be positive if the difference favors method 1, and negative if the difference favors method 2.

[6] Although these latter data are not presented in CHAPTER 3, they are in the detailed tables of the results which are available from the library of the Center for the Advanced Study of Educational Administration, University of Oregon, Eugene, Oregon 97403, on a cost basis.

On the other hand, if a distribution turns out to be bell-shaped, we will conclude that the variation in the resulting distribution is probably caused by measurement errors and any form of control analysis would only reduce the observed variation and the remaining results would tend to cluster about the center of the distribution more than in the original.

This latter effect was demonstrated in a recent review by Stickell (1963) of televised and face-to-face instruction. In this review, 250 comparisons of ETV and conventional face-to-face instruction from 31 research reports were classified as interpretable, partially interpretable or uninterpretable depending upon the adequacy of the experimental design and statistical analysis employed in making each comparison. The results are summarized in TABLE 2. As can be seen, no-difference was found in 73% of the uninterpretable comparisons, 87% of the partially interpret-

Table 2. A Summary of the Results of 250 Comparisons of ETV and Face-to-Face Instruction According to the Adequacy of the Experimental Design and Statistical Analysis Employed in Making Each Comparison. (Stickell, 1963)

	The Comparisons			
The Results Were:	Favored Face-to-Face Instruction	Showed No Difference	Favored ETV	Totals
Interpretable	0 (0%)	10 (100%)	0 (0%)	10
Partially interpretable	0 (0%)	20 (87%)	3 (13%)	23
Uninterpretable	28 (13%)	158 (73%)	31 (14%)	217
Total sample	28 (13%)	188 (75%)	34 (14%)	250

able comparisons and 100% of the interpretable comparisons of the academic achievement of students exposed to ETV and students exposed to face-to-face instruction. The significant point is that by eliminating the uninterpretable and partially interpretable comparisons from his sample, Stickell did not find totally new results—he simply reduced the variation or error which was inherent in his total sample.

The next step in the statistical analysis involved computing statistics to accompany each of the observed distributions. A mean t-ratio was computed by adding the separate t-ratios and dividing the total by the number of comparisons which were represented, symbolically,

$$\text{MEAN} = T = \frac{\sum_{i=1}^{n} t_i}{N}$$

and a standard deviation was computed by dividing the sums of squares by the number of comparisons less one degree of freedom which was "used up" in computing the mean.

$$\text{STANDARD DEVIATION} := \text{SD} = \frac{\sum_{i=1}^{n} (t_i - T)^2}{N - 1}$$

Finally, a difference-of-means test was computed for each type of comparison using the following test statistic,

$$t = \frac{T - \mu}{\frac{s}{\sqrt{N-1}}}$$

$$\text{where } \mu = 0$$

in order that we might get a rough idea of the degree to which the average difference in group performances (T) of each type of comparison actually differed from a "true" difference of zero.

DISCUSSION

With the analyses outlined above, we felt that our initial goals had been achieved. We not only had a sign test of all differences measured which represented the greatest number of available comparisons, but we also had a test of standardized differences in measured group performance which represented a more precise measure of the differences.

Reducing Redundancy

Computation of the measures discussed above for the greatest number of possible comparisons revealed the consistent finding of no difference between any two methods of teaching.

However, these results contain a great deal of redundancy. That is, in many cases we have examined the same result more than once. For example, in study number 006, we examined comparative scores for the same groups on 5 different measures: essay work, quizzes, a mid-term examination, a final examination and semester averages. The differences reported for these 5 measures are probably not independent, and by including the results of the comparisons of the two groups on each of these measures we have probably achieved the same result 5 times. The effect of including measures which are non-independent is to obscure the "true" value of the differences over a number of comparisons, as well as to give

an inflated impression of the total number of comparisons actually being examined. With this in mind, we proceeded to eliminate redundant results from our sample. The criteria used to eliminate comparisons were: (1) only one comparison of any two groups was to be included in the sample; (2) mean-gain scores had precedence over post-test scores; (3) comparisons made on the basis of final examinations had precedence over other comparisons; and, (4) if mean-gain scores or post-test scores were not presented, the comparisons utilized were to be chosen using a random technique.

After the redundant results had been eliminated from the sample, the statistical analyses discussed above were repeated. The results of these analyses are presented in the histogram and detailed statistics of each figure and also in APPENDIX B. Because the independent results are not subject to possible distortion by redundancy, we have based the analysis of CHAPTER 3 only upon them.

APPENDIX
B

Additional Data on Comparative Analysis of College Teaching Methods

Additional data were available for some comparisons among teaching methods not presented in the text. In the following table the data for such additional comparisons are set forth. These additional results accord with the conclusions reached in our main discussion.

Table 1.
Additional Data on Comparative Analysis of College Teaching Methods

Methods Compared		
Method 1	Method 2	N
Lecture	Lecture 3/wk + Discussion 1/wk	1
Lecture	Lecture 2/wk + Discussion 1/wk	2
Lecture	Lecture + Discussion (50/50)	4
Lecture	Lecture 1/wk + Discussion 2/wk	1
Discussion	Lecture 2/wk + Discussion 1/wk	2
Discussion	Lecture + Discussion (50/50)	8
Discussion	Lecture 1/wk + Discussion 2/wk	2
Lecture	Tutorials	3
Lecture	Lecture + Supervised Independent Study	1
Discussion	Tutorials	2
Supervised Independent Study	Lecture 3/wk + Discussion 1/wk	2
Supervised Independent Study	Lecture 2/wk + Discussion 1/wk	2
Supervised Independent Study	Lecture + Discussion (50/50)	19
Self-Study	Lecture	20
Autonomous Small Groups	Lecture	1
Self-Study	Discussion	2
Autonomous Small Groups	Discussion	1
Self-Study	Lecture + Discussion	1
Self-Study	Supervised Independent Study	11
Autonomous Small Groups	Supervised Independent Study	1

Signed Differences			Standardized Differences				
% Favoring Method 1	% Favoring Method 2	% Showing No Difference	N	Mean of Differences	SD of Differences	t	P
0.0%	100.0%	0.0%			Insufficient Data		
100.0	0.0	0.0			Insufficient Data		
50.0	25.0	25.0	3	—0.29	0.37	1.08	.50 > P > .40
0.0	100.0	0.0			Insufficient Data		
50.0	0.0	50.0	2	—1.08	1.53	0.71	P > .50
50.0	50.0	0.0	8	0.41	1.63	0.67	P > .50
0.0	50.0	50.0	2	0.60	0.85	0.71	P > .50
66.7	33.3	0.0	2	—0.32	2.10	0.15	P > .50
0.0	100.0	0.0			Insufficient Data		
100.0	0.0	0.0			Insufficient Data		
0.0	100.0	0.0	2	0.89	0.20	4.50	.20 > P > .10
0.0	100.0	0.0			Insufficient Data		
52.6	47.4	0.0	15	0.09	1.07	0.32	P > .50
60.0	40.0	0.0	20	—0.44	1.12	1.72	.20 > P > .10
0.0	100.0	0.0			Insufficient Data		
100.0	0.0	0.0	2	—0.88	0.01	123.90	.01 > P > .001
100.0	0.0	0.0			Insufficient Data		
0.0	100.0	0.0			Insufficient Data		
27.3	72.7	0.0	11	0.21	0.82	0.82	.50 > P > .40
0.0	100.0	0.0			Insufficient Data		

Bibliography

STUDIES FULLY UTILIZED

001 Asch, M. J., "Nondirective Teaching in Psychology: An Experimental Study," *Psychological Monographs*, 45:1-24. (1951).

050 Balcziak, L. W., "The Role of the Laboratory and Demonstration in College Physical Science in Achieving the Objectives of General Education," Doctoral Dissertation, University of Minnesota, 1953.

032 Bane, C. L., "The Lecture Vs. the Class-Discussion Method of College Teaching," *School and Society*, 21:300-302. (1925)

002 Barnard, W. H., "Note on the Comparative Efficacy of Lecture and Socialized Recitation Versus Group Study Method," *Journal of Education*, 27:388-390. (1936)

003 Barnes, D. G. and Douglass, H. R., "The Value of Extra Quiz Sections in the Teaching of History," Unviersity of Oregon Publications, Educational Series I, 1:276-284. (1929)

004 Beach, L. R., "Sociability and Academic Achievement in Various Types of Learning Situations," *Journal of Educational Psychology*, 51:208-212. (1960)

033 Bechtel, L. P., "Comparative Effects of Differentiated Teaching Methods on Certain Personality Characteristics of College Students," Doctoral Dissertation, New York University, 1963.

052 Bradley, R. L., "Lecture-Demonstration Versus Individual-Laboratory Work in a Natural Science Course at Michigan State University," Doctoral Dissertation, Michigan State University, 1962.

034 Burke, H. R., "An Experimental Study of Teaching Methods in a College Freshman Orientation Course," Doctoral Dissertation, Boston University, 1955.

053 Caro, P. W., Jr., "The Effect of Class Attendance and 'Time Structured' Content on Achievement in General Psychology," *Journal of Educational Psychology*, 53:76-80. (1962)

055 Churchill, R. D., "Evaluation of Independent Study in College Courses," Doctoral Dissertation, University of Minnesota, 1960.

056 Corey, S. M., "Learning from Lectures Versus Learning from Reading," *Journal of Educational Psychology*, 25:459-470. (1934)

057 Coulson, J. E. and Silberman, H. F., "Effects of Three Variables in a Teaching Machine," *Journal of Educational Psychology*, 51: 135-143. (1960)

058 Davis, D. R. and Douglass, H. R., "The Relative Effectiveness of Lecture-Recitation and Supervised Individual Methods in the Teaching of Unified Mathematics in College," *University of Oregon Publications, Educational Series I*, 1:313-316. (1929)

059 Dearden, D. M., "An Evaluation of the Laboratory and Supplementary Teaching Techniques Used in a College General Biology Course," Doctoral Dissertation, University of Minnesota, 1959.

060 Dobyns, R. A., "An Experiment with Programmed Instruction in Teaching College Algebra," Doctoral Dissertation, George Peabody College for Teachers, 1963.

036 Fitzgerald, A. I., "A Study of the Relative Effectiveness of Selected Instructional Procedures in a College Course in Children's Literature," Doctoral Dissertation, University of Missouri, 1960.

061 Flynn, J. T., "The Influence of Programmed Instruction Upon Learning in Educational Psychology," Doctoral Dissertation, Indiana University, 1963.

062 Fodor, J. T., "A Comparative Study of Two Approaches to Health Instruction at the College Level," Doctoral Dissertation, University of Calfornia, Los Angeles, 1963.

008 Gerberich, J. R. and Warner, K. O., "Relative Instructional Efficiencies of the Lecture and Discussion Methods in a University Course in American National Government," *Journal of Educational Research*, 29:574-579. (1936)

037 Gore, A. E., "Individualized Instruction Through Team Learning in a College Course in General Psychology," Doctoral Dissertation, Boston University, 1962.

009 Griffen, K. and Bowers, J. W., "An Experimental Study of the Use of Lectures to Large Groups of Students in Teaching the Fundamentals of Speech," *Journal of Educational Research*, 55:383-385. (1962)

011 Haigh, B. V. and Schmidt, W. H., "Learning of Subject Matter in Teacher-Centered and Group-Centered Classes," *Journal of Educational Psychology*, 47:295-301. (1956)

064 Hartman, G. W., "Comparative Pupil Gains Under Individual Conference and Classroom Instruction," *Journal of Educational Psychology*, 26:367-372. (1935)

039 Hatcher, B. E., "An Experimental Study of Three Different Patterns of Student Participation in a General Education Science Course for College Freshmen," Doctoral Dissertation, Wayne State University, 1957.

066 Hough, J. B., "An Analysis of the Efficiency and Effectiveness of Selected Aspects of Machine Instruction," *Journal of Educational Research*, 55:467-471. (1962)

040 Jenkins, R. L., "The Relative Effectiveness of Two Methods of Teaching Written and Spoken English," Doctoral Dissertation, Michigan State College, 1951.

067 Jensen, B. T., "Instruction and Personality as Factors in Student Performance," *Journal of Educational Research*, 47:529-535. (1954)

068 Kellems, R. L., "A Comparative Analysis of the Effect of the Use of a Programmed Text on Achievement and Efficiency in College Algebra," Doctoral Dissertation, Indiana University, 1964.

069 Key, J. F., "An Investigation in the Teaching of Quadratic Equations Using Programmed Instruction," Doctoral Dissertation, George Peabody College for Teachers, 1964.

071 Larson, H. A., "An Evaluation of Laboratory Teaching Methods for a Beginning College Course in Principles of Accounting," Doctoral Dissertation, University of Minnesota, 1962.

017 Leton, D. A., "An Evaluation of Course Methods in Teaching Child Development," *Journal of Educational Research*, 55:118-122. (1961)

018 Lifson, N., Rempel, P., and Johnson, J. A., "A Comparison Between Lecture and Conference Methods of Teaching Physiology," *Journal of Medical Education*, 31:376-382. (1956)

072 Little, C. E., "An Experimental Study of Programmed Instruction in College Algebra at Colorado State College," Doctoral Dissertation, Colorado State College, 1964.

073 Maiser, T., "A Study of the Effectiveness of a Programmed Unit on the Hysteria Syndrome in Three Colleges," Doctoral Dissertation, Temple University, 1962.

075 Marr, J. N., *et al.*, "The Contribution of the Lecture to College Teaching," *Journal of Educational Psychology*, 51:277-284. (1960)

076 McKeegan, H. F., "Assessment of a Graduate Level Self-Teaching Technique Based on a Combination of Programmed Instruction and the Case Method," Doctoral Dissertation, University of Pittsburgh, 1963.

043 Montean, J. J., "An Experimental Study of the Use of Discussion in General Chemistry and General Science as a Means of Increasing Group Growth in Critical Thinking," Doctoral Dissertation, Syracuse University, 1959.

020 Morris, V. P. and Douglass, H. R., "The Relative Effectiveness of the Problem and Lecture Methods of Instruction in Principles of Economics," *University of Oregon Publications, Educational Series I*, 1:285-292. (1929)

021 Olson, K. V., "An Experimental Evaluation of a Student-Centered Method and a Teacher-Centered Method of Biology Science Instruction for General Education of College Students," *Journal of Experimental Education*, 27:225-230. (1959)

077 Payne, V. F., "The Lecture-Demonstration and Individual Laboratory Methods Compared. III. Experimental," *Journal of Chemical Education*, 9:1277-1294. (1932)

044 Rafter, C. D., "A Comparison of the Relative Effectiveness of Two Methods of Teaching a Course in Physical Science to Sophomore College Students," Doctoral Dissertation, New York University, 1956.

078 Rea, R. G., "Comparison of Programmed Instruction with the Traditional Method of Instruction in Parliamentary Procedure at Auburn University," Doctoral Dissertation, Auburn University, 1965.

079 Richards, A. A., "The Development of a Linear Programmed Instructional Unit in Educational Administration for the Study of the Theory of Decision-Making and an Evaluation of its Usefulness as a Teaching Device," Doctoral Dissertation, New York University, 1965.

080 Robson, C. J., "An Experimental Study of the Efficacy of Two Instructional Methods in Introductory College Psychology," Doctoral Dissertation, University of Minnesota, 1961.

045 Slomowitz, M., "A Comparison of Personality Changes and Content Achievement Gains Occurring in Two Modes of Instruction," Doctoral Dissertation, New York University, 1955.

046 Smith, J. P., "A Study of Outcomes of Instruction in a Psychology Class with Special Reference to Two Teaching Methods," Doctoral Dissertation, Ohio State University, 1954.

084 Smith, N. H., "The Teaching of Elementary Statistics by the Conventional Classroom Method of Instruction Versus the Method of Programmed Instruction," Doctoral Dissertation, University of Colorado, 1963.

025 Spence, R. B., "Lecture and Class Discussion in Teaching Educational Psychology," *Journal of Educational Psychology*, 19:454-462. (1928)

086 Taylor, H. R., "An Experiment with Independent Study," *University of Oregon Publications, Educational Series I*, 1:300-312. (1929)

026 Tuttle, H. S. and Douglass, H. R., "Project Teaching in a College Course in Educational Psychology," *University of Oregon Publications, Educational Series I*, 1:293-299. (1929)

027 Ward, J. M., "Group-Study Versus Lecture-Demonstration Method in Physical Science Instruction for General Education Students," *Journal of Experimental Education*, 24:199-210. (1956)

088 Wardian, J.F., "An Experiment Concerning the Effectiveness of Programed Learning for Use in Teaching the Fundamentals of Music," Doctoral Dissertation, Washington State University, 1963.

028 White, J. R., "A Comparison of the Group-Laboratory and the Lecture-Demonstration Methods in Engineering Instruction," *Journal of Engineering Education*, 36:50-54. (1945)

030 Wieder, G. S., "Group Procedures Modifying Attitudes of Prejudice in the College Classroom," *Journal of Educational Psychology*, 45:332-344. (1954)

048 Weinick, G. D., "The Comparative Effectiveness of Two Teaching Methods in Attaining Specific Course Objectives," Doctoral Dissertation, New York University, 1960.

STUDIES PARTIALLY UTILIZED
(Sign Test Only of Differences in Outcome)

049 Baldwin, R. E., "The Development and Evaluation of Voluntary-Choice Programmed Instruction in Mathematics," Doctoral Dissertation, University of Minnesota, 1964.

093 Becker, Samuel L,. *et al.*, "Teaching by the Discussion Method: An Experiment with Television Discussion, Small Group Discussion, Large Group Discussion, and the Lecture Method," State University of Iowa, 1958. (Mimeo)

035 Delong, A. R., "The Relative Effectiveness of Two Methods of Teaching Social Science at the College Level," Doctoral Dissertation, University of Michigan, 1949.

006 Edmonson, J. B. and Mulder, F. J. "Size of Class as a Factor in University Instruction," *Journal of Educational Research*, 9:1-12. (1924)

010 Guetzkow, H., Kelly, L. E., and McKeachie, W. J., "An Experimental Comparison of Recitation, Discussion and Tutorial Methods in College Teaching," *Journal of Educational Psychology*, 45:193-207. (1954)

065 Heckman, M. A., "The Relative Merits of Two Methodologies for Teaching Verbal Arithmetic Problems to Undergraduate Elementary Education Majors," Doctoral Dissertation, Indiana University, 1962.

012 Hurst, J. H., "The Relationship Between Teaching Methods and Course Objectives in Educational Psychology," *Journal of Educational Research*, 57:147-151. (1963)

014 Johnson, D. M. and Smith, H. C., "Democratic Leadership in the College Classroom," *Psychological Monographs*, 67:1-20. (1953)

041 Kantasewi, N., "Experiments in the Use of Programmed Materials in Teaching an Introductory Course in the Biological Sciences at the College Level," Doctoral Dissertation, Pennsylvania State University, 1964.

019 Longstaff, H. P., "Analysis of Some Factors Conditioning Learning in General Psychology," *Journal of Applied Psychology*, 16:9-48; 131-166. (1932)

022 Rasmussen, G. R., "An Evaluation of a Student-Centered and Instructor-Centered Method of Conducting a Graduate Course in Education," *Journal of Educational Psychology*, 47:449-461. (1956)

082 Scheidemann, N. W., "A Comparison of Two Methods of College Instruction," *School and Society*, 25:672-674. (1927)

083 Skogman, K. W., "Conventional Versus Programmed Instruction in English Grammar at Dana College," Doctoral Dissertation, University of Nebraska Teachers College, 1964.

024 Smith, H. C., "Teamwork in the College Class," *Journal of Educational Psychology*, 46:274-286. (1955)

132 Stuit, Dewey B., *et al.*, "Iowa Closed-Circuit Television Teaching Experiment: Summary Report," State University of Iowa, 1956. (Mimeo)

087 Thomas, J. M. and Fattu, N., "A Comparison of Laboratory and Non-Laboratory Instruction in Freshman English," *Studies in Higher Education, Biennial Report of the Committee on Educational Research, 1938-1940*, University of Minnesota. (1941)

029 Wispe, L. G., "Evaluating Section Teaching Methods in the Introductory Course," *Journal of Educational Research*, 45:161-186. (1951)

031 Zelleny, L., "Experimental Appraisal of a Group Learning Plan," *Journal of Educational Research*, 34:37-42. (1940)

STUDIES EXAMINED BUT NOT UTILIZED
(No Data Presented on Which to Base Sign Test or Other Analyses)

051 Bartz, W. H. and Darby, C. L., "A Study of Supervised and Non-supervised Programmed Instruction in the University Setting," *Journal of Educational Research*, 58:208-211. (1965)

005 Casey, J. E. and Weaver, B. E., "An Evaluation of Lecture Method and Small Group Method of Teaching in Terms of Knowledge of Content, Teacher Attitude, and Social Status," *Journal of Colorado-Wyoming Academy of Science*, 4:54. (1956)

054 Christmas, E. P., "An Experimental Evaluation of Programmed Instruction in Teaching Agronomy at the College Level, Doctoral Dissertation, Purdue University, 1964.

007 Eglash, A., "A Group-Discussion Method of Teaching Psychology," *Journal of Educational Psychology*, 45:257-267. (1954)

038 Gorman, Rev. R. M., "An Experimental Study of the Effectiveness of Several Combinations of Lecture and Discussion for Students in an Educational Psychology Course," Doctoral Dissertation, Fordham University, 1962.

063 Greene, E. B., "The Relative Effectiveness of Lecture and Individual Reading as Methods of College Teaching," *Genetic Psychology Monographs*, 4:463-563. (1928)

013 Husband, R. W., "A Statistical Comparison of the Efficacy of Large Lecture Versus Smaller Recitation Sections Upon Achievement in General Psychology," *Journal of Psychology*, 31:297-300. (1951)

070 Kruglak, H., "Experimental Outcomes of Laboratory Instructions in Elementary College Physics," *American Journal of Physics*, 20:136-141. (1952)

015 Krumboltz, J. D. and Farquar, W. W., "The Effect of Three Teaching Methods on Achievement and Motivational Outcomes in a How-To-Study Course," *Psychological Monographs*, 71:1-26. (1957)

016 Lahti, A. M., "The Inductive-Deductive Method and the Physical Science Laboratory," *Journal of Experimental Education*, 24: 149-163. (1956)

042 Larson, R. W., "The Differential Effect of Two Methods of College Teaching on the Attitudes of Students in a Pre-Service Education Program," Doctoral Dissertation, University of Minnesota, 1956.

074 Marcuse, F. L., "On Methods of Teaching Elementary Psychology," *Journal of Educational Psychology*, 42:236-240. (1951)

023 Ruja, R., "Outcomes of Lecture and Discussion Procedures in Three College Courses," *Journal of Experimental Education*, 22:385-394. (1954)

081 Ryan, T. A., "Testing Instructional Approaches for Increased Learning," *Phi Delta Kappan*, 46:534-536. (1965)

085 Stone, J. B., "The Effects of Learner Characteristics on Performance in Programmed Text and Conventional Text Formats," Doctoral Dissertation, University of Utah, 1963.

047 Timmel, G. B., "A Study of the Relationship Between Methods of Teaching a College Course in Mental Hygiene and Changes in Student Adjustment Status," Doctoral Dissertation, Cornell Univ., 1954.

089 Whitney, C. B., "The Effect of General College Physics Laboratory Instruction on Selected Student Non-Academic Characteristics," Doctoral Dissertation, University of Kansas, 1965.

REVIEWS OF RESEARCH

1938 Monroe, W. and Marks, A., "General Methods of Teaching," *Educational Administration and Supervision*, 24:497-512, 581-592.

1939 Underwood, F. M., "Evolution of Educational Method," *Educational Method*, 18:230-233.

1942 Wolfe, D., "The First Course in Psychology," *Psychological Bulletin*, 39:685-712.

1952 Ruja, H., "Experimenting with Discussion in College Teaching: A survey of recent research," *Educational Administration and Supervision*, 39:321-342.

1954 McKeachie, W., "Student-Centered Versus Instructor-Centered Instruction," *Journal of Educational Psychology*, 45:143-150.

1955 Axelrod, J., "Group Dynamics, Nondirective Therapy, and College Teaching," *Journal of Higher Education*, 26:200-207.

1955 Birney, R. and McKeachie, W. J., "The Teaching of Psychology: A Survey of Research Since 1942," *Psychological Bulletin*, 52: 51-68.

1956 Buxton, C. E., COLLEGE TEACHING: A PSYCHOLOGIST'S VIEW. New York: Harcourt, Brace & Co., Inc.

1958 Bureau of Institutional Research, "A Review of the Literature Concerning Studies of College Teaching Methods and Class Size," University of Minnesota.

1960 Wingo, M., "Methods of Teaching," in C. HARRIS, Ed.), ENCYCLOPEDIA OF EDUCATIONAL RESEARCH, 3rd Ed., New York: Macmillan.

1962 McKeachie, W. J., "Procedures and Techniques of Teaching: A Survey of Experimental Studies," in N. Sanford, THE AMERICAN COLLEGE, New York: John Wiley and Sons, Inc.

1963 Felder, B. D., "Characteristics of Independent Study Practices in Colleges and Universities of the United States," Doctoral Dissertation, University of Texas.

1963 McKeachie, W. J., "Research on Teaching at the College and University Level," in N. Gage (Ed.), HANDBOOK OF RESEARCH ON TEACHING, Chicago: Rand McNally & Co., 1118-1172.

1963 Wallen, N. and Travers, R., "Analysis and Investigation of Teaching Methods," in N. Gage, (Ed.), HANDBOOK OF RESEARCH ON TEACHING, Chicago: Rand McNally & Co., 448-505.

1964 Powell, J., "Experimentation and Teaching in Higher Education," *Educational Research*, 6:179-191.

1965 McKeachie, W. J., "The Discussion Group," Center for Research on Learning and Teaching, Memo to the Faculty, No. 14.

1968 Taveggia, Thomas C., "Differences Between the Lecture and Group Discussion Methods of Teaching at the College Level," Master's Thesis, University of Oregon.